Cognition 258 (2025) 106083

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

People accurately predict the shape but not the parameters of skill learning curves

Xiuyuan Zhang [a,*], Samuel D. McDougle [a,b], Julia A. Leonard [a,b]

[a] Department of Psychology, Yale University, United States of America
[b] Wu Tsai Institute, Yale University, United States of America

ARTICLE INFO

Keywords:
Skill learning
Learning curves
Metacognition
Affect

ABSTRACT

Decades of research have shown that skill learning often unfolds exponentially — people improve rapidly early on, and then performance gradually levels off. Given how important expectations of learning are for actual learning, we explored whether people accurately intuit this canonical time course of skill learning. Across six preregistered experiments ($n = 500$), we find that people correctly predict that skill learning curves (error reductions over time) on a novel visuomotor task will follow an exponential decay function, both for an imagined naïve player and for themselves, before engaging with the task. Moreover, people are sensitive to conditions that merit exponential learning within a bounded time frame and only predict these curves when an imagined player puts in effort and the task is not too difficult. However, people systematically misestimate specific parameters of skill learning (e.g., initial and average performance, and rate of improvement), which relates to reduced affect at the beginning of learning. Critically, these negative effects can be ameliorated by practice: Providing people with minimal practice reduces their prediction errors and, in turn, buffers them from negative feelings at the beginning of learning.

1. Introduction

Learning curves have been a cornerstone of the psychological lexicon for centuries: Educators, coaches, employers, and psychophysicists often talk about "steep" or "plateauing" learning curves to describe the experience of improving at a task. Functionally, learning curves represent performance over time and capture the rate at which a new skill is acquired. Although there is well-established research on how people track their progress over time *during* learning (e.g., people choose to restudy items they got wrong when preparing for a test, see Kornell & Metcalfe, 2006; Payne, Youngcourt, & Beaubien, 2007; Ten, Kaushik, Oudeyer, & Gottlieb, 2021; Simon & Bjork, 2001), less is known about whether or how people represent learning curves *prior* to learning.

Understanding how people think about learning curves before learning is important as these expectations might shape which tasks individuals choose to pursue and whether they decide to persist versus quit in the face of challenges. Consider the desire of many people to take up an instrument, like violin or piano, during the grueling early days of the 2020 COVID-19 quarantine. Given the choice among many instruments, people may have gravitated toward an instrument that they thought would be easy to pick up, rather than one that they thought would take years to grasp. This is in line with prior work showing that people typically spend their time on tasks where they experience steep learning curves (Ten et al., 2021). Furthermore, if someone expects swift progress learning the piano, but it turns out to be slow, then they may develop negative feelings and prematurely quit (Rutledge, Skandali, Dayan, & Dolan, 2014; Dai, Dietvorst, Tuckfield, Milkman, & Schweitzer, 2018). To our knowledge it is unclear whether people accurately intuit how their future learning will progress across time prior to the experience of learning.

Decades of research have documented that learning curves on motor and cognitive tasks usually proceed according to a decelerating exponential function (Thorndike, 1913; Newell & Rosenbloom, 2013; Heathcote, Brown, & Mewhort, 2000; Moskowitz, Gale, Gallivan, Wolpert, & Flanagan, 2020): Novices will experience substantial initial improvement over a short period of time, but the amount of improvement per unit of time decreases as performance asymptotes (Heathcote et al., 2000; Krakauer, Hadjiosif, Xu, Wong, & Haith, 2019). Although different individuals can approach these asymptotes at different rates, the general shape of learning across individuals is relatively stereotyped. Given that learning usually unfolds following this canonical pattern, it seems plausible that people may have mental representations of learning

* Corresponding author at: address: Department of Psychology, Yale University, 100 College Street, New Haven, CT 06511, United States of America.
E-mail addresses: flora.zhang@yale.edu (X. Zhang), samuel.mcdougle@yale.edu (S.D. McDougle), julia.leonard@yale.edu (J.A. Leonard).

https://doi.org/10.1016/j.cognition.2025.106083
Received 12 July 2024; Received in revised form 31 January 2025; Accepted 6 February 2025
Available online 20 February 2025

that also follow this exponential decay function.

Prior work in the “Judgments of Learning (JOL)” literature has shown that people can monitor and evaluate their own performance at specific time points *during* learning and strategically allocate more time studying items that are not “too hard” nor “too easy” before a test (Rhodes, 2016; Bjork, Dunlosky, & Kornell, 2013; Finn & Metcalfe, 2008; Koriat & Bjork, 2006; Arbuckle & Cuddy, 1969; Koriat, 1997). However, relatively few studies have asked individuals to predict their future performance across multiple trials *prior* to learning. Studies that have taken this approach find a “stability bias”: People predict that their memory for word pairs will remain the same across four rounds of studying and testing (Koriat, Bjork, Sheffer, & Bar, 2004; Kornell & Bjork, 2009; Kornell, Rhodes, Castel, & Tauber, 2011; though follow-up studies suggested that this result may be an artifact of framing, Ariel, Hines, & Hertzog, 2014). This line of research has also primarily focused on memory tasks and only probed predictions of learning concerning the very beginning of the task, where there is usually rapid improvement. Thus, it is unknown whether people predict that their rate of improvement will eventually decline with practice, especially on novel skill tasks. Importantly, this question is best suited for tasks where ceiling (or near-ceiling) performance takes a long time and, to avoid framing effects, where few linguistic cues are employed.

Our central hypothesis is that people intuitively think of skill learning as rapid improvement followed by a slow decline in progress (consistent with the canonical exponential decay learning curves). However, there are two reasons to believe that people may struggle with accurately predicting the shape of learning curves. First, prior work shows that people have an exponential growth bias, where they tend to “linearize” observations and underestimate exponential growth over time. This bias has been identified in a wide variety of domains from predicting duckweed growth rates in a pond (Wagenaar & Sagaria, 1975; Wagenaar & Timmers, 1978) to estimating personal savings and investments in financial decision-making (McKenzie & Liersch, 2011). Thus, people might similarly predict a more linear function for learning curves.

Second, people tend to be especially overly optimistic about their future self (Koriat et al, 2004; Koriat, Lichtenstein, & Fischhoff, 1980; Koriat & Bjork, 2005; Lefebvre, Lebreton, Meyniel, Bourgeois-Gironde and Palminteri, 2017; Garrett and Sharot, 2017; Horn & Loewenstein, 2024). Indeed, prior work shows that people systematically over-predict their starting performance and under-predict their rate of growth on the first few trials of a memory task (Ariel et al., 2014). It is only with practice that people dampen and better calibrate their expectations (Finn & Metcalfe, 2007; Finn & Metcalfe, 2008; Koriat, Sheffer, & Ma’ayan, 2002). Thus, it is possible that people’s over-optimism about their performance may lead them to predict better overall performance and, in turn, more linear improvement.

If people can construct a mental model of future learning trajectories akin to exponential decay learning curves, how might they do this? Do they use a simple heuristic that exponential learning curves underlie all learning? Or instead, do people base their predictions of learning curves on features of the learner and the specific task. Prior work suggests that a learner’s performance on a task is influenced by both task features and the learner’s own attributes (Anderson, Lohse, Lopes, & Williams, 2021; Duckworth, Eichstaedt, & Ungar, 2015; Guadagnoli & Lee, 2004; Pashler, McDaniel, Rohrer, & Bjork, 2008). Unsurprisingly, learners start off with worse performance and make less progress over time on harder compared to easier tasks (Ahissar & Hochstein, 1997; Odic, Hock, & Halberda, 2014), and learners who put in less effort will not learn as much as someone who puts in more effort (Frömer, Stürmer, & Sommer, 2016; Lee, Swinnen, & Serrien, 1994). Thus, if people are making judgments about learning curves based on task and learner features, given a bounded time frame, they may only predict exponential learning curves when the task is not too hard and the player puts in effort. In cases where the player does not try hard and/or the task is very difficult, people may instead predict worse average performance and minimal gains as the player is presumably far from reaching a performance asymptote in a bounded time frame. Since prior work has shown that people are sensitive to both task difficulty and effort when making learning curve judgments and when evaluating someone’s knowledge state or performance (Berke, Tenenbaum, Sterling, & Jara-Ettinger, 2023; Heller, Arnold, Klein, & Tanenhaus, 2015; Hodges & Lohse, 2020; Muradoglu & Cimpian, 2020; Song & Schwarz, 2008), it seems plausible that they might also rely on inferences about task difficulty and player effort when estimating future learning curves.

In sum, we drew inspiration from the aforementioned work on learning curves, metacognition, and affective science to ask how humans construct and represent the time course of future skill learning. To do this, we needed to create a task that is (1) novel to participants, (2) has an exponential learning curve where people donot hit ceiling performance within the allotted time of the experiment, (3) has minimal linguistic demands to avoid framing effects (see Ariel et al., 2014), and (4) can capture intuitive predictions of trial-by-trial performance.

To fulfill these criteria, we created a novel visuomotor skill learning task called “Lolli-toss”, where the goal is to gain points by launching lollipops toward the center of a target. This computer-based task requires participants to learn the function of two keys — one that can stop a lollipop from sliding sideways on the bottom of the screen with a click and one that can toss the lollipop toward a target based on the length of the press. Although the goal of Lolli-toss is similar to other games, such as throwing darts, the functionalities of playing are unique to this game and thus novel to all players. Furthermore, reaching optimal performance on Lolli-toss takes time, more than the 50 trials allotted.

Lolli-toss was also specifically designed to capture precise trial-by-trial predictions of future learning curves: Instead of having people choose from different learning trajectories or draw out the shape of the learning curve as a graph (which is notoriously hard for people; Ciccione, Sablé-Meyer, & Dehaene, 2022; Melnik-Leroy et al., 2023), we had participants simply report where on the target they thought the lollipop would land after different amounts of experience. This captures predictions of performance over time (also known as a learning curve; distinct from learning outcomes; Heathcote et al., 2000; Solum, Lorås, & Pedersen, 2020; Horn & Loewenstein, 2024). The benefit of measuring predictions of learning curves in the motor domain, rather than the cognitive domain, is that motor tasks often allow for more overt, direct predictions (e.g., instead of asking “how many novel-word pairs will you remember on this trial”, participants can simply click where they think their lollipop will land). Participants’ predictions, as well as where the lollipops actually landed during subsequent learning, allowed us to measure both predicted and actual learning curves using the same metric (namely, Euclidean distance from the target’s center).

In Experiments 1a, 1b, and 1c, we first establish that the learning curve on our novel visuomotor task generally follows an exponential decay function, and then examine whether people accurately predict that a naïve player’s learning curve on this task would show rapid improvement followed by smaller gains (consistent with an exponential decay function), using two distinct prediction methods. In Experiment 2, we test whether people’s predictions of learning curves are informed by perceived task difficulty and player effort. Finally, in Experiments 3a and 3b, we explore individuals’ metacognitive awareness of their *own* learning curves, and how (in)accurate expectations about learning curves have consequences for motivation and affect during actual learning. Material, analysis code, and data from all six experiments can be accessed on the OSF repository: https://osf.io/xzm5c/, and all scripts and regression tables are reported in the Supplemental Information (SI). Taken together, these experiments aim to uncover how humans build and represent the progression of skill learning before embarking on acquiring a new skill.

2. Experiment 1a

We first asked people to play our novel visuomotor learning task, Lolli-toss, to establish that learning follows the canonical motor learning exponential decay function. This experiment was preregistered: https://osf.io/cw7g2.

2.1. Methods

2.1.1. Participants

We recruited 50 adult U.S. participants ($M_{\text{age}} = 28.52$, $SD_{\text{age}} = 9.93$; 74 % female, 26 % male) online through Prolific. The self-reported racial and ethnic makeup of participants was White (50 %), Hispanic or Latino (18 %), Asian (16 %), Black or African American (10 %), and multiracial or biracial (6 %). Forty-eight percent of participants reported having a bachelor's degree or higher. Based on preregistered exclusion criteria, five additional participants were excluded for not having a negatively signed learning rate (extracted using a linear model $error = rate * trial + constant$; we consider this a relaxed criteria, since participants were not excluded based on the magnitude or the statistical significance of the learning rate parameter), which we took as a sign of participant inattention or lack of effort on the task. Note that our main results replicate when we include these five excluded participants (see SI).

2.1.2. Stimuli

We designed a novel online visuomotor game called Lolli-toss. The goal of Lolli-toss is to get as many points as possible by "tossing" lollipops that move back and forth along the bottom of the screen toward the bullseye of a target board above (Fig. 1a). The lollipops always appeared first on a randomized side (left or right) of the window before moving horizontally across the bottom of the screen. Players used two keys to play the game: The "space" bar stopped the lollipop, and the "enter" key launched the lollipop forward toward the target. The amount of time a player held down the "enter" key corresponded to the vertical distance that the lollipop traversed (e.g., the longer they held down the "enter" key, the farther the lollipop went). We set the optimal interval to 1406.67 ms, with hold intervals less than 940.33 ms or more than 1873 ms fully missing the target. If a player got the lollipop in the bullseye of the target, they received 50 points. For each concentric ring outside of the bullseye (starting from red, ending at white), they could get 30, 20, 10, and 5 points respectively, and 0 points for landing the lollipop outside of the target. A version of Lolli-toss can be accessed on the OSF repository.

2.1.3. Procedure

The task consisted of a training phase and a Play phase. In the training phase, participants were introduced to the goal, features (e.g., moving lollipop), and scoring system of the novel Lolli-toss game. Participants were then told how the "space" bar and the "enter" key controlled the lollipops and were able to press the keys to confirm that these keys were functional on their keyboard without the target board or lollipop present (to prevent explicit learning on the board). Participants had to pass two comprehension questions before proceeding further. In the Play phase, participants played 10 rounds of Lolli-toss, with 5 tosses per round (to reduce fatigue), for a total of 50 tosses (trials). After each toss, participants saw their toss score and their total score. In between each round, participants were shown their score from the previous round, their cumulative score, and a reminder of how many rounds they had completed.

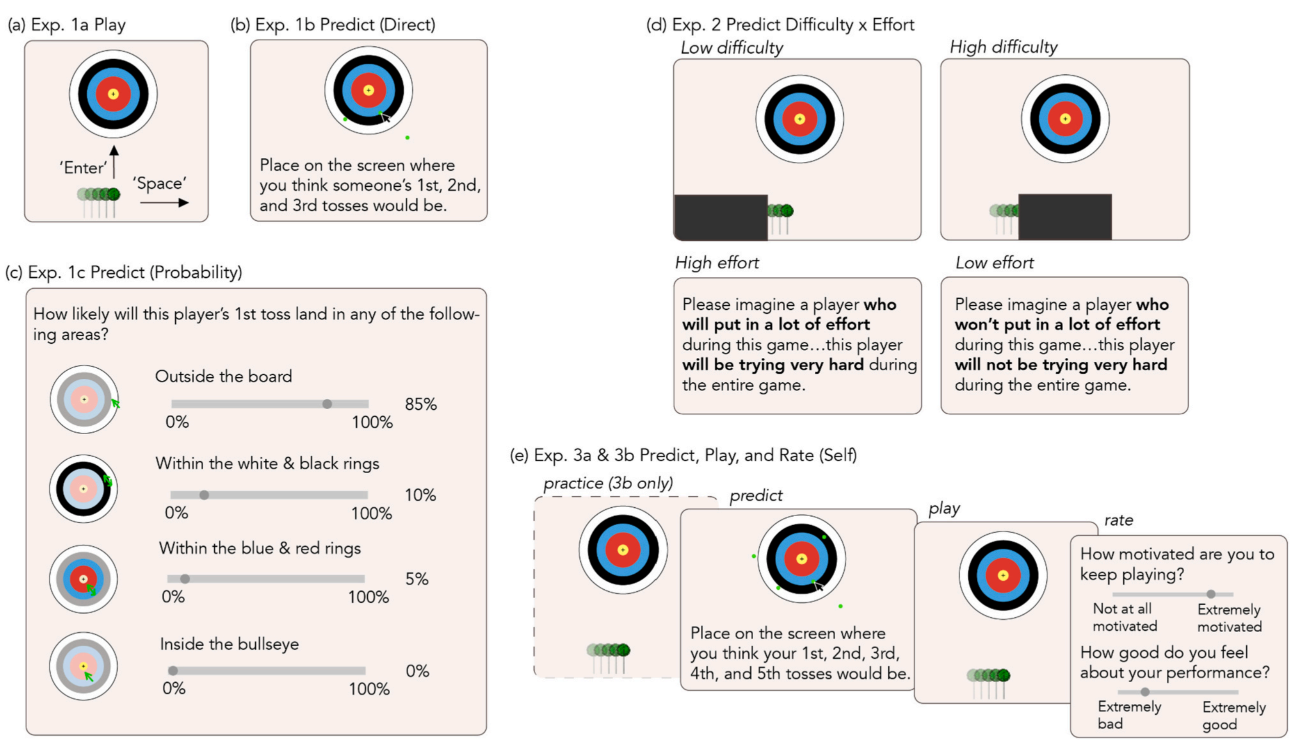

Fig. 1. Schematics of Experiments 1a, 1b, 1c, 2, 3a and 3b. (a) In Experiment 1a, participants played an online game, Lolli-toss, for 50 tosses. The goal of Lolli-toss is to launch a lollipop that moves from side to side into the yellow center of the target by first stopping its path with the "space" key and then tossing it forward with the "enter" key. The distance that the lollipop moves upward is proportional to the duration that the "enter" key is held down. (b) In Experiment 1b, participants made predictions for three consecutive trials by clicking on the screen where they predicted a beginner player's lollipops would land every seven trials. They also predicted where a participant's last toss would land on the 50th trial. (c) In Experiment 1c, participants judged the likelihood of a beginner player's lollipop landing in each of the four areas using sliders for the same trials as Experiment 1b. (d) In Experiment 2, participants predicted performances for four conditions (2 difficulty × 2 effort levels) using the same method as Experiment 1b. (e) In Experiment 3a, participants first predicted their own performance in Lolli-toss for 50 trials, and then played the game for 50 trials and rated their motivation and feelings in between each round (after every 5 trials). Experiment 3b provided participants with three chances to practice Lolli-toss before making predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Additional post-task questions on subjective task difficulty, prior gaming experience, and demographic information were collected at the end of the game (see SI for measures and associated analyses).

2.2. Results

2.2.1. Preregistered analyses

To quantify participants' actual and predicted trial-by-trial performance, we calculated the Euclidean distance from the center of the lollipop head to the center of the target (i.e., their error for a given toss) and fit each participant's actual or predicted learning curve with two models: an exponential decay function ($Error = a \exp(-b*Trial) + c$), and a linear function ($Error = -b*Trial + c$). We chose the exponential decay function because (1) it clearly describes our hypothesis that people intuit that performance quickly improves early on and then slowly asymptotes, (2) is a canonical function used in prior motor learning literature to describe learning curves (Heathcote et al., 2000), and (3) would imply, as we predict, that the rate at which performance improves is proportional to the current level of performance. Based on work showing that people tend to linearize exponential functions (Hutzler et al., 2021; McKenzie & Liersch, 2011; Wagenaar & Sagaria, 1975), we chose the linear function as our null hypothesis comparison point: If our hypothesis is correct, people's predicted and actual learning curves should better fit the exponential compared to the linear function.

As expected, learning curves on Lolli-toss were better characterized by an exponential decay function than a linear function. Because the exponential model contained an additional free parameter relative to the linear model, we evaluated model performance for all participants using the Akaike information criterion (AIC), which penalizes for the number of free parameters in a given model, and where a lower AIC score indicates better fit (Akaike, 1974). A paired Wilcoxon signed-rank test revealed that the exponential model (*Mean* = 555.62, *Median* = 551.18, *SD* = 37.92) outperformed the linear model (*Mean* = 566.02, *Median* = 562.92, *SD* = 37.34) and had significantly lower AIC scores across the group ($V = 53$, $p < .001$; Fig. 2a). Additional exploratory analyses and discussion using logarithmic and power decay functions to fit participant data can be found in the SI.

2.3. Interim discussion

In line with prior work on sensorimotor skill learning (Heathcote et al., 2000), Experiment 1a established that learning curves on our novel skill task, Lolli-toss, was best fit by an exponential decay model rather than a linear function.

3. Experiments 1b - 1c

We next explored whether people can accurately predict the exponential shape of learning curves in Lolli-toss. In Experiment 1b, we asked people to simulate a naïve learner's trial-by-trial performance and make direct predictions about the location of a learner's toss on the game screen. To test the robustness of our findings, in Experiment 1c, we queried predictions using an additional paradigm commonly used in the JOL literature: asking participants to estimate the likelihood of a toss landing in specific predefined areas on the game screen (see examples in Koriat et al., 2004; Kornell & Bjork, 2009; Rhodes, 2016). Both experiments were preregistered (Experiment 1b: https://osf.io/4dhf2; 1c: https://osf.io/q6ytr).

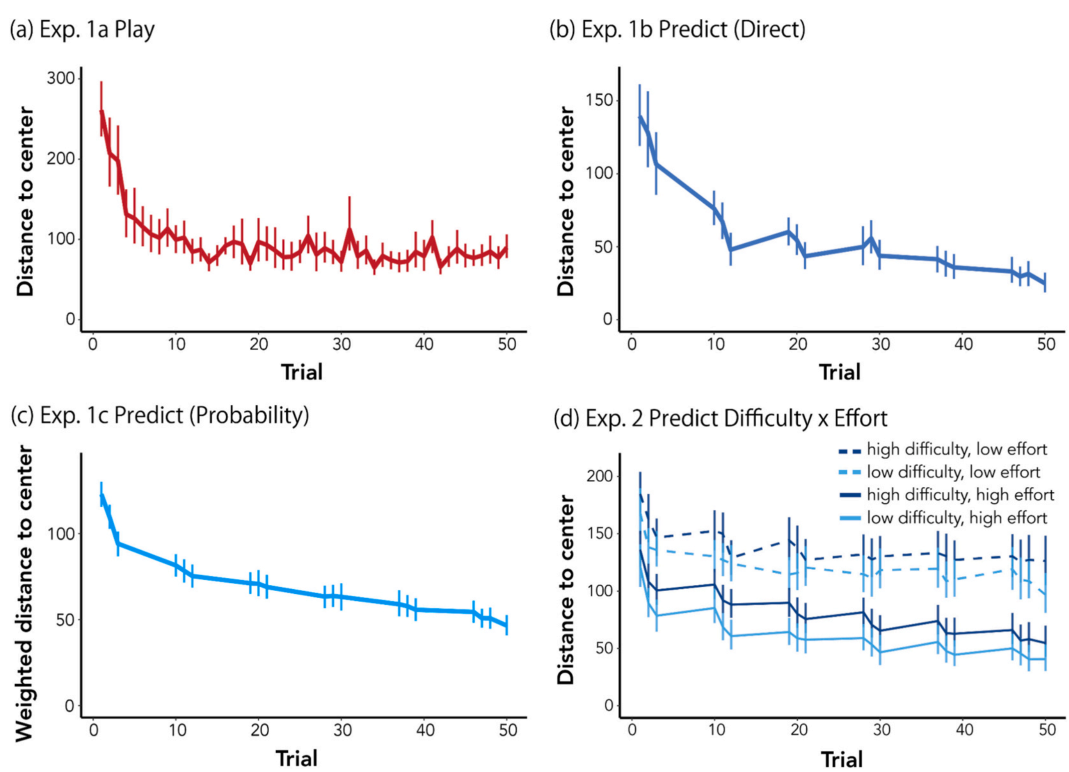

Fig. 2. Average actual and predicted learning trajectories in Experiments 1a, 1b, 1c and 2. (a) Actual learning trajectory in Experiment 1a. (b) Predicted learning trajectory for a naïve player in Experiment 1b via trial-by-trial direct clicks of where lollipops will land. (c) Predicted learning trajectory for a naïve player in Experiment 1c via trial-by-trial probability estimates. (d) Predicted learning trajectories for four conditions in Experiment 2 that varied player effort (high vs. low) and task difficulty (high vs. low). Dark blue indicates high difficulty and light blue indicates low difficulty; solid lines indicate high effort while dotted lines indicate low effort. Note that error bars indicate 95 % bootstrapped confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1. Methods

3.1.1. Participants

In Experiment 1b, we recruited 50 adult U.S. participants (M_{age} = 32.24, SD_{age} = 8.89; 52 % female, 44 % male, 2 % non-binary, 2 % preferred not to answer) online through Prolific. The self-reported racial and ethnic background of the sample was White (74 %), Asian (14 %), Black or African American (6 %), and multiracial or biracial (6 %). The majority of participants (74 %) reported having a bachelor's degree or above. Four additional participants were excluded for not predicting improvement (i.e., a negatively signed learning rate) across trials based on the preregistered criteria. This criterion was intended to exclude participants who may be responding less attentively or randomly. However, note that our main results replicated when including these participants excluded for not predicting improvement in Experiments 1b and 1c (see SI for details). In Experiment 1c, we recruited an identical sample size of 50 adult U.S. participants (M_{age} = 27.54, SD_{age} = 7.58; 76 % female, 22 % male, and 2 % non-binary) on the same online platform. The self-reported racial and ethnic background of the sample was White (76 %), Asian (8 %), Hispanic or Latino (8 %), Black or African American (6 %), and multiracial or biracial (2 %). Half of the participants reported to have a bachelor's degree or higher. An additional 13 participants were excluded for not predicting any improvement across trials based on the same preregistered criteria as Experiment 1a.

3.1.2. Stimuli

We developed a prediction version of Lolli-toss with the same features in Experiments 1b and 1c (Fig. 1b and c) as the play version of Lolli-toss in Experiment 1a.

3.1.3. Procedure

The task in Experiments 1b and 1c consisted of a training phase and a Predict phase. In the training phase, participants were introduced to Lolli-toss as in Experiment 1a. However, instead of being asked to play Lolli-toss, participants were instructed that they would predict a beginner player's progress in the game. They were told that this player had never seen Lolli-toss before and would be playing Lolli-toss for 50 trials. Participants were given one opportunity to experience how the "space" and "enter" keys work to move the lollipop without the presence of the target board to avoid direct learning. To proceed further, participants had to pass three comprehension questions about the goal of the task in both experiments, and two additional comprehension questions about likelihood estimations in Experiment 1c (see SI for details of comprehension questions).

In the Predict phase, participants were asked to make trial-by-trial predictions of the lollipop landing locations. To reduce task redundancy and fatigue, participants predicted a total of 19 out of 50 trials across seven rounds. Participants made predictions on the first three consecutive trials per round for the first six rounds. This resulted in 7-trial intervals between rounds (e.g., they predicted trials 1, 2, 3 in round 1 and trials 10, 11, 12 in round 2, etc.). To ensure a matched ending trial between the prediction responses and the learning responses from Experiment 1a, participants also predicted the landing location of the last toss (i.e., the 50th trial) in the seventh round. In Experiment 1b, participants were instructed to make predictions by clicking directly on the game screen ("You have 3 clicks in this round. Place on the screen where you think someone's 1$^{\text{st}}$, 2$^{\text{nd}}$, and 3$^{\text{rd}}$ tosses would be."). Participants' previous clicks were visible to them during a given round with a green dot to reduce memory demands (see Fig. 1b) and were removed when a new round started. In Experiment 1c, participants were asked to estimate the likelihood of the toss landing on each of four possible regions (e.g., "If you have to make your best guess, how likely will this player's 1$^{\text{st}}$ throw land in any of the following areas?"; see Fig. 1c). Participants responded with four sliders (0 % - 100 %), one for each region. This approach allowed us to track changes in the relative likelihood distribution for tosses landing in each of the four regions across trials. Participants were only able to proceed to the next trial if the sum of all four sliders was 100 %.

After completion of the prediction phase, participants were asked the same set of post-study questions as in Experiment 1a (see SI).

3.2. Results

3.2.1. Preregistered analyses

As hypothesized, participants predicted that learning curves on Lolli-toss would follow an exponential decay function in Experiments 1b and 1c. Experiment 1b used the same analysis approach as in Experiment 1a, where participants' predicted trial-by-trial performance was quantified using the Euclidean distance and fit to both exponential ($Error = a$ exp.$(-b{*}Trial) + c$) and linear function ($Error = -b{*}Trial + c$) models. A paired Wilcoxon signed-rank test revealed that the exponential models ($Mean = 179.01$, $Median = 180.97$, $SD = 21.18$) outperformed the linear models ($Mean = 188.79$, $Median = 187.21$, $SD = 18.06$) with significantly lower AIC scores ($V = 174$, $p < .001$ for all participants; Fig. 2b). Experiment 1c used participants' weighted trial-by-trial predictions — converting the reported probabilities to Euclidean distance in the task space and computing weighted distances for participants using the likelihood estimation on each trial — and found that the exponential models ($Mean = 130.23$, $Median = 130.73$, $SD = 20.61$) outperformed the linear models ($Mean = 146.26$, $Median = 146.52$, $SD = 20.27$), and had significantly lower AIC scores ($V = 106$, $p < .001$ for all participants; Fig. 2c; see SI for details on probability-to-distance conversion).

3.2.2. Exploratory analyses

It is possible that the exponential shape of participants' predicted learning curves was artificially derived by people predicting two linear processes: Initial linear improvement followed by a linear flat performance curve at some plateau. To rule out this possibility, we also fit participants' predicted learning trajectories in Experiment 1b using segmented regression models: Participants' predicted trajectories were fit by two linear models with a breaking point. If participants predicted rapid improvement early on followed by slower progress later during learning (an exponential decay curve), we would expect to see participants predicting performance gains in both segments. Alternatively, if participants predicted initial improvement followed by constant performance, we would expect participants to predict flat learning rates (i.e., not significantly different than 0) for the second segment. We found evidence for the former: Participants predicted that a beginner player would improve quickly in the first segment ($Mean_{slope\ 1} = -28.21$, $Median_{slope\ 1} = -11.52$, $SD_{slope\ 1} = 38.51$) compared to the second segment ($M_{slope\ 2} = -0.48$, $Median_{slope\ 2} = -0.64$, $SD_{slope\ 2} = 4.62$; paired Wilcoxon signed-rank test $V = 105$, $p < .001$) and participants still predicted improvement during the second segment rather than constant performance (one-sample Wilcoxon signed-rank test $V = 232$, $p < .001$). See SI for additional model comparisons and discussion using logarithmic and power decay functions to fit participants' predicted learning curves.

To examine if participants were accurate at predicting the specific parameters of a beginner player's learning trajectory, we next compared participants' predictions in Experiment 1b to the actual performance data in Experiment 1a. We found that participants were more optimistic about a beginner player's performance compared to people's actual performance in the task. Participants in Experiment 1b predicted that beginner players would make smaller errors across trials (average of performance across all trials; $Mean_{predict} = 58.29$, $Median_{predict} = 55.15$, $SD_{predict} = 26.27$; $Mean_{learn} = 94.05$, $Median_{learn} = 91.68$, $SD_{learn} = 27.18$; Wilcoxon rank-sum test $W = 381$, $p_{FDR\text{-}corrected} < .001$), have better performance at the start (calculated by averaging performance across trials 1–3; $Mean_{predict} = 114.85$, $Median_{predict} = 106.41$, $SD_{predict} = 52.59$; $Mean_{learn} = 222.09$, $Median_{learn} = 201.79$, $SD_{learn} = 108.25$; $W = 388$, $p_{FDR\text{-}corrected} < .001$), and have better final performance (calculated by averaging performance across the last three matched trials, trials 47, 48, and 50; $Mean_{predict}$ = 27.20, $Median_{predict}$ = 19.16, $SD_{predict}$ = 20.68; $Mean_{learn}$ = 82.35, $Median_{learn}$ = 75.59, SD_{learn} = 31.63; W = 149, $p_{FDR\text{-}corrected} < .001$) than they actually do. Participants also predicted players to have less variability in the size of their errors compared to players' actual error variability in the task ($Mean_{predict}$ = 41.93, $Median_{predict}$ = 36.50, $SD_{predict}$ = 18.60; $Mean_{learn}$ = 74.57, $Median_{learn}$ = 66.01, SD_{learn} = 29.67; W = 376, $p_{FDR\text{-}corrected} < .001$). Lastly, participants predicted that players would improve at the task at a slower rate than they actually did ($Mean_{predict}$ = −0.28, $Median_{predict}$ = −0.12, $SD_{predict}$ = 0.34; $Mean_{learn}$ = −0.42, $Median_{learn}$ = −0.36, SD_{learn} = 0.32; W = 1704, $p_{FDR\text{-}corrected}$ = .002).

3.3. Interim discussion

We found converging evidence across methods that people accurately predict the exponential decay shape of learning on a novel visuomotor task prior to any direct task experience. When asked to reason about a naïve player's learning curve by making either point estimates (Experiment 1b) or likelihood estimates (Experiment 1c) of performance, people's predictions were better fit by an exponential decay model than a linear model. However, a comparison between Experiments 1a and 1b revealed that people also misestimate the specific parameters of learning curves, leading to overly optimistic predictions of a naïve player's overall performance, but an underestimation of their learning rate. In Experiment 2, we turn our attention to examining whether information about task difficulty and player effort impacts people's predictions about future learning curves.

4. Experiment 2

In Experiment 2, we explore whether people always predict exponential skill learning curves given a bounded time frame, or instead, construct and adjust their predictions in response to both task and player features. Specifically, we probed people's learning curve predictions using a within-subjects 2 × 2 design, manipulating both task difficulty and player effort (see Fig. 1d). We varied perceived task difficulty by controlling whether a visual occluder appears in the center of the lollipop's path, blocking the desired lollipop toss location, or to the left of the screen. We manipulated player effort by explicitly telling participants to predict performance for one player who puts in a lot of effort during the game and one who does not.

We predict that participants will modify their predicted learning curves in response to both features (Experiment 2 is preregistered: https://osf.io/nj8kh). Specifically, we hypothesize that participants will predict better performance (average, starting, and ending), as well as a steeper learning curve, when the player puts in more effort and the task is not difficult. In the case where the player puts in little effort and the task is difficult, we hypothesize that participants might predict flatter learning curves in the same period of 50 trials. Thus, it is possible that only when the player puts in high effort and the task is not difficult do participants predict learning curves that follow an exponential decay function.

4.1. Methods

4.1.1. Participants

We recruited 100 adult U.S. participants (M_{age} = 35.29; SD_{age} = 11.22; 51 % female, 45 % male, 3 % non-binary, and 1 % other) through Prolific. The self-reported racial and ethnic background of the sample was White (80 %), Asian (8 %), Black or African American (5 %), other (5 %), Native Hawaiian or other Pacific Islander (1 %), and prefer not to answer (1 %). Half of the sample held a bachelor's degree or higher. An additional 29 participants were excluded from further analyses based on preregistered exclusion criteria (24 participants were excluded due to incorrectly answering the perceived difficulty question; 5 participants were excluded due to incorrectly answering one or more comprehension check questions about player effort; see script in SI for exact wording). Because we did not expect that participants would predict negative learning rates in all conditions (e.g., when a *low effort* player is playing a *high difficulty* game), we did not preregister any exclusion criterion based on the direction of the slope (i.e., predicted learning rates) as prior experiments.

4.1.2. Stimuli

To manipulate task difficulty, we designed two variations of the original Lolli-toss game by adding a visual occluder (80 pixels × 392 pixels) to the moving lollipop's horizontal path at different positions (Fig. 1d). In the *high difficulty* version, the occluder blocked the center of the horizontal lollipop path, thereby blocking visual access to the main lollipop release location. In the *low difficulty* version, the occluder blocked the left side of the path, leaving the optimal lollipop release location in sight. Participants were not explicitly told the difficulty level for each version and instead had to infer it from visual features, as is often the case in the real world (indeed adults and children accurately judge task difficulty from static features of a task, Gweon, Asaba, & Bennett-Pierre, 2017; Yildirim et al., 2019). To manipulate player effort, participants learned about a novice player who will either put in a lot of effort and try very hard (*high effort* condition), or not put in a lot of effort nor try very hard (*low effort* condition) during the entire game.

4.1.3. Procedure

As in previous experiments, this task consisted of a training phase and a Predict phase. In the training phase, participants were introduced to the game of Lolli-toss and a naïve player who will play the game for 50 trials. Additionally, participants were presented with information about the two versions of the task and two types of players who put in more or less effort before proceeding to complete the same comprehension questions as in Experiments 1b and 1c.

In the Predict phase, participants were assigned to all four conditions in a randomized order (a within-person 2 × 2 design crossing difficulty and effort). For each condition, participants were shown the task version and player effort information before proceeding to make predictions about a player's performance. Participants completed 19 predictions on the same trials as Experiments 1b and 1c. After making predictions, participants were asked a forced-choice attention check question about the effort level of the player they just predicted. As preregistered, if participants did not correctly answer all four effort questions, they were excluded from further analyses.

Upon completion of all predictions, participants answered a series of post-task questions (see SI for details and additional analyses). Since task difficulty was never explicitly introduced to participants during the experiment, they were asked two comprehension questions about difficulty, one for each Lolli-toss version. We preregistered including participants who correctly rated the center occluder version as more difficult than the side occluder version. If participants answered that both were of equal difficulty or that the *low difficulty* version was more difficult than the *high difficulty* version, they were excluded.

4.2. Results

4.2.1. Preregistered analyses

Participants' predictions of learning trajectories on Lolli-toss were impacted by both player effort and task difficulty (Fig. 2d). As in prior experiments, we calculated the Euclidean distance across participants' 19 predictions in each condition to create measures of trial-by-trial performance and learning curves (performance over trials). Three linear mixed-effects models with fixed effects of effort and difficulty and random intercepts by participant revealed that participants predicted that greater effort and lower task difficulty would lead to better average performance (e.g., smaller errors; player effort: b = −58.39, $p < .001$; task difficulty: b = −18.27, $p < .001$), better starting performance (player effort: b = − 50.34, $p < .001$; task difficulty: b = −17.85, $p <$

.001), and better ending performance (player effort: $b = -66.29$, $p < .001$; task difficulty: $b = -17.74$, $p < .001$). Interaction models between player effort and task difficulty were all not significant (see SI for full regression tables).

Notably, only learning curves from the *high effort*, *low difficulty* condition were better fit by an exponential model (*Mean* = 173.17, *Median* = 174.79, *SD* = 21.42) than a linear model (*Mean* = 178.32, *Median* = 178.63, *SD* = 19.89; $V = 1698$, $p = .03$ for 95 converged participants; $p > .4$ for the other three conditions; see SI for details and additional replication analyses using data from the first condition only). For the participants whose exponential models converged across conditions ($n = 67$), we ran linear mixed-effects models predicting learning rates (extracted from the exponential models) with fixed effects of task difficulty and player effort and random intercepts by participants. Results showed significant effects of difficulty ($b = -0.09$, $p = .02$) and effort ($b = 0.08$, $p < .05$) on predicted learning rates. Participants judged that learning an easier task would result in a faster learning rate compared to learning a more difficult task. Moreover, participants judged that a player who puts in more effort would have a slower learning rate. Further analyses with an additional interaction term revealed a significant interaction between effort and difficulty ($b = -0.19$, $p = .02$). Interrogating this interaction revealed that learning rates did not differ by task difficulty when a player did not try (low difficulty: *Mean* = −0.29, *Median* = −0.08, *SD* = 0.38; high difficulty: *Mean* = −0.29, *Median* = −0.11, *SD* = 0.37; $V = 944$, $p_{FDR\text{-}corrected} = .8$). However, when a player put in high effort, learning rates were steeper on the easier task (*Mean* = −0.30, *Median* = −0.07, *SD* = 0.38) compared to the harder task (*Mean* = −0.12, *Median* = −0.04, *SD* = 0.21, $V = 1652$, $p_{FDR\text{-}corrected} = .003$; see SI for all pairwise comparisons). We considered these results suggestive due to a smaller sample of converged participants.

4.3. Interim discussion

Experiment 2 revealed that people only expect skill learning to unfold exponentially when an imagined player puts in effort and the task is not very difficult in a bounded time frame. People also expect that performance will be on average better when a player puts in more effort and the task is not very difficult. These predictions are in line with prior work showing that effort increases learning outcomes and task difficulty decreases learning outcomes (Metcalfe & Kornell, 2005; Pashler et al., 2008). Thus, people may use prior beliefs about task difficulty and player effort in conjunction with explicit information about these features to intuitively construct expected learning curves. In Experiment 3a, we shift our focus to examining how individuals predict their own future learning curves.

5. Experiment 3a

Prior experiments all focused on third party predictions of learning curves. An open question is whether people similarly predict that their own learning curve will follow an exponential decay function, as well as whether they have good metacognitive awareness of features of their unique expected learning curve on a novel task. Going beyond past work that has looked at people's moment-to-moment metacognitive judgments of learning for only a few discrete time points after some experience (Kornell & Metcalfe, 2006; Kornell & Bjork, 2009), in Experiment 3a we asked whether people can predict the shape and features of their future learning curves over time with minimal task exposure. Specifically, we asked participants to first make predictions about their own performance on Lolli-toss over 50 trials and then play the game for 50 trials.

Importantly, perceptions of one's own learning trajectory may affect what tasks people choose to learn, as well as whether they persist in learning. Indeed, prior work has shown that expecting swift success and then facing early setbacks causes premature quitting (Dai et al., 2018). One possibility is that a large prediction error between expected performance and actual performance causes people to feel unhappy (Rutledge et al., 2014) and perhaps give up. However, it is also possible that poor performance alone, rather than the discrepancy between predicted and actual performance, leads to negative affect or lower motivation. As people's memory of their predicted learning curves may fade or be hard to track during actual learning, it is possible that actual performance may be more salient than predicted performance for motivational and affective processes as learning progresses. Thus, in Experiments 3a, we test these different possibilities by measuring people's affective and motivational responses to their performance on Lolli-toss after they predicted their own performance on the same game. Experiment 3a is preregistered: https://osf.io/fxas2.

5.1. Methods

5.1.1. Participants

We recruited 125 adult U.S. participants ($M_{age} = 37.04$; $SD_{age} = 12.46$; 51.2 % female, 44.8 % male, and 4 % non-binary) through Prolific. The self-reported racial and ethnic background of the sample was White (77.6 %), Black or African American (9.6 %), Asian (6.4 %), other (4.8 %), and American Indian or Alaska Native (1.6 %). About half of the sample (48.8 %) held a bachelor's degree or higher. An additional 46 participants were excluded from further analyses based on preregistered exclusion criteria (3 participants did not answer the training comprehension check questions correctly after three tries, 3 participants failed the writing comprehension check, and, using the same preregistered criteria as Experiments 1a-1c, 40 participants did not predict or show improvement; however, see SI for details showing that main results replicate when including these 40 participants).

5.1.2. Stimuli

To collect participants' predictions of their own performance and their actual performance, we combined the stimuli used in Experiments 1a and 1b.

5.1.3. Procedure

This study consisted of a training phase, a Predict phase, and a Play phase (Fig. 1e). In the training phase, participants were told that they would predict their own performance on a novel game called Lolli-toss. They were then introduced to how the game works as in Experiment 1a and asked to answer three comprehension check questions about the goal of the task. If a participant failed to respond to the check questions correctly after three tries, they were directed to exit the experiment and received partial compensation.

The Predict phase was identical to Experiment 1b except for the following changes. Unlike Experiment 1b, participants made predictions for all 50 trials to gain more precise estimates for analyses comparing predicted and actual learning curves. Participants predicted their lollipop landing locations in 10 rounds, with five predictions per round.

The procedure for the Play phase was identical to Experiment 1a with two additional exploratory questions about participants' motivational and affective states in between each round. Before the first round, participants were asked to rate their initial motivation level and their affective state on two sliders. Then, after playing each round, participants were asked to rate their momentary motivation level and affective state. After completing the 50 tosses, participants were asked a set of post-task questions (including two open-ended comprehension questions; see SI for details and additional analyses).

5.2. Results

5.2.1. Preregistered analyses

As in prior experiments, participants' predicted and actual learning curves on Lolli-toss were best fit by an exponential decay function (Fig. 3a). Following the same analyses as in Experiment 1b, Wilcoxon

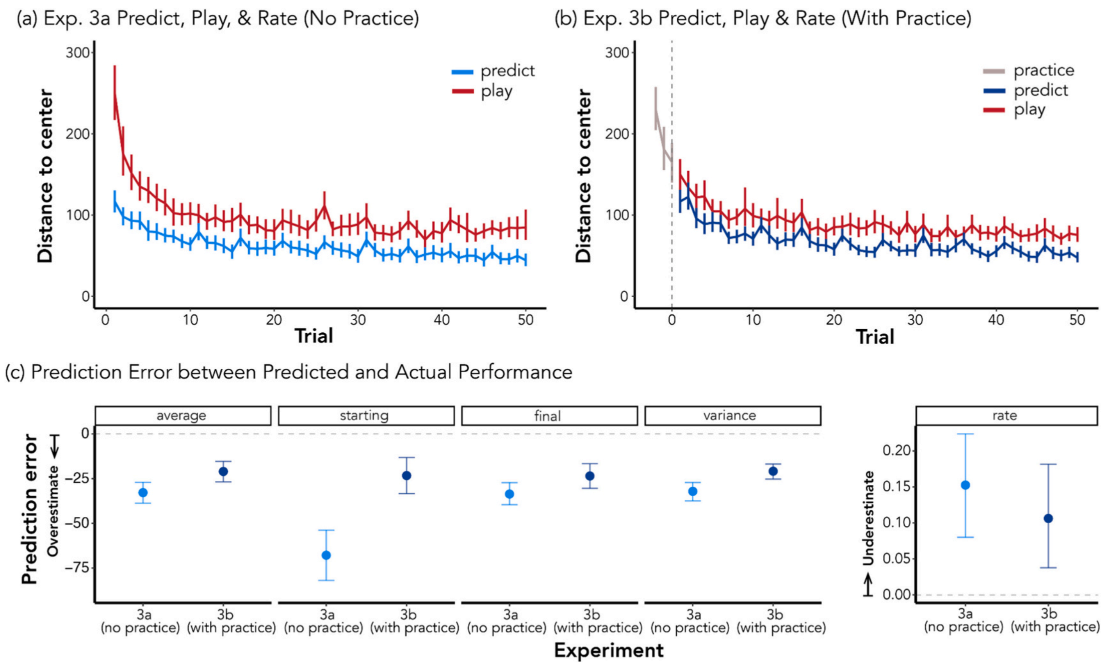

Fig. 3. Experiments 3a and 3b predicted and actual learning trajectories. (a) First-person predicted (in light blue) and actual (in red) learning trajectories in Experiment 3a. (b) First-person predicted (in navy) and actual (in red) learning trajectories in Experiment 3b, where participants were allowed to practice Lolli-toss prior to making any predictions. (c) Prediction errors (the difference between predicted and actual performance) in Experiment 3a (in light blue) and Experiment 3b (in navy) for average, starting, and final performance, variance, and learning rates. All error bars indicate 95 % bootstrapped confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

signed-rank tests revealed that participants' predicted learning curves were better described by the exponential model (*Mean* = 474.09, *Median* = 476.23, *SD* = 46.20) than the linear model (*Mean* = 482.08, *Median* = 479.41, *SD* = 41.94; $V = 2440$, $p < .001$ for all but one participant whose exponential model did not converge). Following the same analyses as in Experiment 1a, Wilcoxon signed-rank tests also revealed that participants' actual learning curves were better fit by exponential models (*Mean* = 547.10, *Median* = 549.22, *SD* = 38.97) versus linear models (*Mean* = 555.45, *Median* = 559.82, *SD* = 40.99; $V = 1295$, $p < .001$ for all participants; see SI for additional exploratory analyses and discussion on fitting participant data using logarithmic and power decay functions).

Paired Wilcoxon signed-rank tests revealed that participants were on average overly optimistic about their future learning trajectories, predicting that they would make smaller errors (average of performance across 50 trials; $Mean_{predict} = 58.72$, $Median_{predict} = 54.34$, $SD_{predict} = 28.07$; $Mean_{learn} = 91.54$, $Median_{learn} = 87.52$, $SD_{learn} = 25.36$; $V = 468$, $p_{FDR\text{-}corrected} < .001$), have better starting performance (calculated by averaging performance across trials 1–5; $Mean_{predict} = 93.10$, $Median_{predict} = 87.98$, $SD_{predict} = 42.69$; $Mean_{learn} = 160.95$, $Median_{learn} = 143.76$, $SD_{learn} = 75.33$; $V = 620$, $p_{FDR\text{-}corrected} < .001$), better final performance (calculated by averaging performance.

across trials 45–50; $Mean_{predict} = 44.12$, $Median_{predict} = 40.57$, $SD_{predict} = 28.00$; $Mean_{learn} = 77.74$, $Median_{learn} = 71.18$, $SD_{learn} = 28.92$; $V = 629$, $p_{FDR\text{-}corrected} < .001$), and less variability (standard deviation) in performance ($Mean_{predict} = 33.66$, $Median_{predict} = 30.92$, $SD_{predict} = 12.08$; $Mean_{learn} = 65.78$, $Median_{learn} = 63.32$, $SD_{learn} = 26.49$; $V = 177$, $p_{FDR\text{-}corrected} < .001$) than warranted given their actual performance. On average, participants also thought that they would improve at a slower rate than they actually did ($Mean_{predict} = -0.20$, $Median_{predict} = -0.07$, $SD_{predict} = 0.27$; $Mean_{learn} = -0.35$, $Median_{learn} = -0.22$, $SD_{learn} = 0.33$; $V = 5457$, $p_{FDR\text{-}corrected} < .001$; Fig. 3c).

Participants did not appear to have precise metacognitive predictions of their own specific learning curves: There was a marginally significant correlation between people's predicted average performance and actual average performance ($rho = 0.19$, $p_{FDR\text{-}corrected} = .058$). There was also a similar trend for starting performance ($rho = 0.20$, $p_{FDR\text{-}corrected} = .058$) and final performance ($rho = 0.19$, $p_{FDR\text{-}corrected} = .058$). However, participants did not accurately predict the variance in their performance ($rho = 0.05$, $p_{FDR\text{-}corrected} = .6$) or their learning rates ($rho = 0.11$, $p_{FDR\text{-}corrected} = .3$).

5.2.2. Exploratory analyses

We tested whether people's affect and motivation changed in response to their prediction errors (i.e., the difference between expected and actual performance for each round) and actual performance during the game. Two separate linear mixed-effects models with a fixed effect of round and random intercepts by participants revealed that participants' motivation linearly decreased across rounds of playing Lolli-toss ($b = -1.28$, $p < .001$), but their affect linearly increased across rounds ($b = 0.70$, $p < .001$; see regression tables in SI; Fig. 4a). Due to the high collinearity between participants' prediction errors and actual performance ($rho = -0.60$, $p < .001$), we ran separate linear mixed-effects models with each variable as a fixed effect predicting motivation and affect (while controlling for the fixed effect of round and random intercepts by participants; see SI). When we compared model performance using AIC values, we found that actual performance predicted changes in motivation (performance: $b = -0.03$, $p_{FDR\text{-}corrected} < .001$, $AIC = 10{,}047$; prediction errors: $b = 0.02$, $p_{FDR\text{-}corrected} = .01$, $AIC = 10055$) and affect (performance: $b = -0.17$, $p_{FDR\text{-}corrected} < .001$, $AIC = 10486$; prediction errors: $b = 0.12$, $p_{FDR\text{-}corrected} < .001$, $AIC = 10574$) better than prediction errors across each round. In other words, and unsurprisingly, participants were more motivated and felt better the closer they were to the bullseye. Participants were also more motivated and felt better when they made smaller prediction errors, but these models did not explain as much variance as the performance-only model (see *AICs* above).

As shown in Fig. 4a, we identified qualitatively a sharp decrease in affective judgments between participants' baseline affect judgment to their judgment after completing the first round. Affect rating changes between round 1 and baseline differed significantly from 0 (*Mean* =

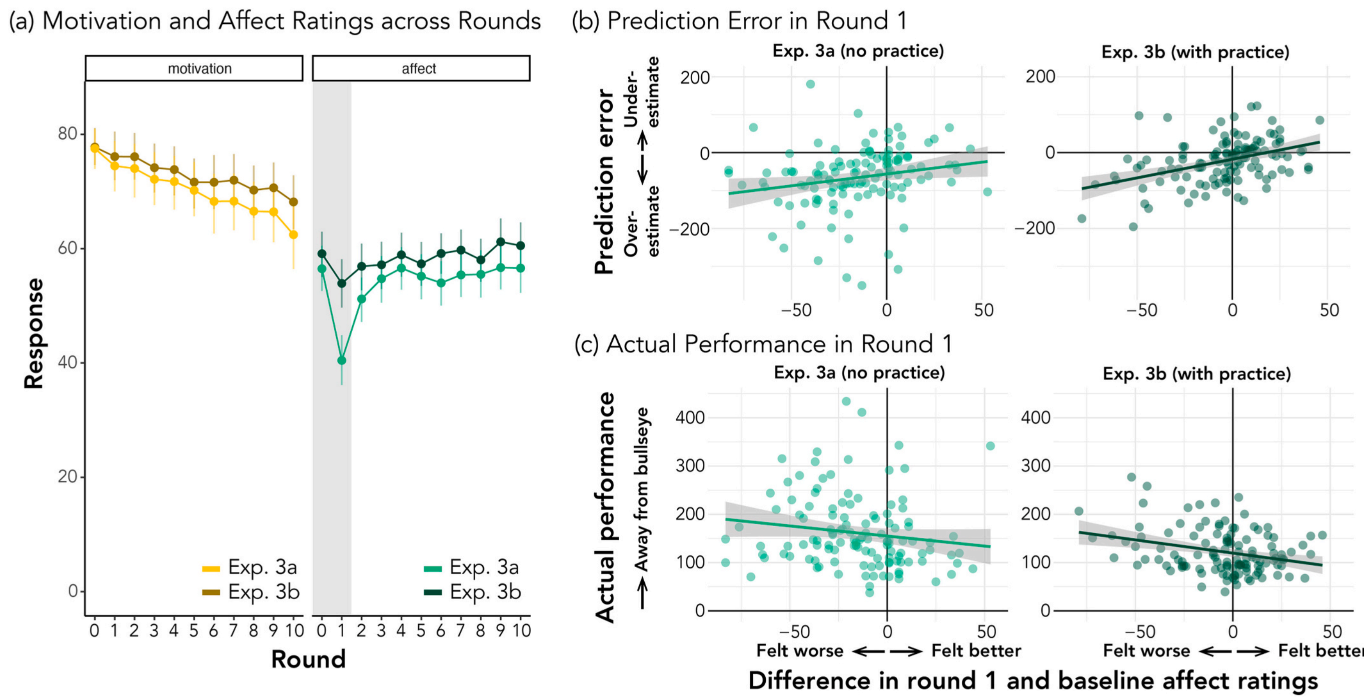

Fig. 4. Momentary motivation and affect ratings during Lolli-toss in Experiments 3a (no practice) and 3b (with practice). (a) Participants' motivation and affect judgments prior to and after each round of playing Lolli-toss. For the motivation panel, average motivation ratings for Experiment 3a are in yellow, and ratings for Experiment 3b are in brown. For the affect panel, average affect ratings for Experiment 3a are in light green, and ratings for Experiment 3b are in dark green. The gray bar highlights the decrease in affect ratings between baseline (round 0) and round 1 in Experiment 3a. The error bars are bootstrapped 95 % confidence intervals. (b) Correlation between prediction errors and the affect difference during round 1. (c) Correlation between actual performance and the affect difference during round 1. Affect difference ratings are calculated by taking the difference between ratings after round 1 and from the baseline. The gray bars in (b) and (c) indicate standard error for the regression line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

−16.04, *Median* = −14, *SD* = 27.36, one-sample Wilcoxon $V = 1449.5$, $p < .001$). This time point also corresponded to one of the largest prediction errors that people made — they thought they would perform significantly better than they actually did at the beginning of the task prior to playing. We theorized that this sudden dip in affective state might be related to the stark difference between what people predict they will do at the beginning of the game and how they actually do. A linear regression predicting changes in affect between the beginning and the end of round 1 using prediction error revealed that the more participants overestimated their performance, the worse they felt about their actual performance ($b = 0.07$, $p_{FDR\text{-}corrected} < .05$, $AIC = 804.26$; Fig. 4b). In contrast, no significant effect was found when predicting changes in affect between the beginning and end of round 1 using actual performance in round 1 ($b = -0.05$, $p_{FDR\text{-}corrected} = .10$, $AIC = 806.67$; see regression tables in SI; Fig. 4c), and this model showed a higher AIC score (worse model performance) compared to the model using only prediction error. That is, performance prediction errors, and not performance alone, appeared to be related to participants' large negative change in affect at the beginning of the game.

Additional exploratory analyses examining the location patterns and learning curves by each component of Lolli-toss — "stopping" on the X-axis and "tossing" on the Y-axis — can be found in SI.

5.3. Interim discussion

Results from Experiment 3a show that people accurately predict the canonical exponential decay shape of their own future skill learning curves on Lolli-toss, but not necessarily the specific features of this process. People are overly optimistic about their performance but underestimate their rate of learning, resulting in miscalibrated predictions of their own learning curves. These inaccuracies may have affective consequences during learning: The larger the prediction error between predicted and actual performance at the beginning of learning, the more negative people felt about their performance. Inspired by the "underconfidence with practice" effect (Koriat et al., 2002; Finn & Metcalfe, 2007; Finn & Metcalfe, 2008), in Experiment 3b, we explored whether limited practice would lower people's prediction errors and in turn reduce the sharp decline in their affect ratings at the beginning of learning.

6. Experiment 3b

In Experiment 3a we found that people tend to overestimate how well they would perform at the beginning of Lolli-toss. This overestimation is associated with feeling worse at the beginning of learning when their expectations do not match their performance. Here we examined whether a brief practice prior to making any predictions would provide people with better metacognitive access to their own performance and, in turn, more accurate predictions of initial performance and better feelings at the beginning of learning. Experiment 3b is preregistered: https://osf.io/gtmhe.

6.1. Methods

6.1.1. Participants

We recruited 125 adult U.S. participants (M_{age} = 35.20; SD_{age} = 11.10; 60.8 % female, 38.4 % male, and 0.8 % non-binary) through Prolific. The self-reported racial and ethnic background of the sample was White (62.4 %), Black or African American (24 %), Asian (7.2 %), other (4 %), Native Hawaiian or other Pacific Islander (1.6 %), and preferred not to answer (0.8 %). Over half of the sample (54.4 %) held a bachelor's degree or higher. An additional 86 participants were excluded from further analyses based on preregistered exclusion criteria (7 participants did not answer the training comprehension check questions correctly after three tries, 12 participants failed the writing comprehension check, and 67 participants did not predict or show improvement; See SI for analyses showing that including these 67 participants did not change main results).

6.1.2. Stimuli

We used the same paradigm as Experiment 3a (Fig. 1e).

6.1.3. Procedure

The procedure is identical to that of Experiment 3a except for the following change: In the training phase, after being introduced to the "stopping" and "tossing" components, participants were given three practice trials to stop the moving lollipop and toss it toward the target (see SI for other minor changes and additional analyses).

6.2. Results

6.2.1. Preregistered analyses

Replicating results from Experiment 3a, participants both predicted their learning trajectories to be exponential and their actual learning curves followed an exponential decay shape (predict AIC values: $Mean_{exp} = 492.12$, $Median_{exp} = 496.78$, $SD_{exp} = 51.93$, $Mean_{linear} = 497.85$, $Median_{linear} = 500.17$, $SD_{linear} = 49.54$, paired Wilcoxon $V = 2858$, $p = .008$ for all participants; practice and learn AIC values: $Mean_{exp} = 586.41$, $Median_{exp} = 586.67$, $SD_{exp} = 41.31$, $Mean_{linear} = 594.25$, $Median_{linear} = 597.36$, $SD_{linear,} = 41.22$, paired Wilcoxon $V = 1184$, $p < .001$ for all but one participant whose exponential model fit did not converge; Fig. 3b).

We next examined if participants over- or under-estimated specific features of their learning trajectories over 50 trials after practice using paired Wilcoxon sign-rank tests. Consistent with results from Experiment 3a (without practice), participants in Experiment 3b overestimated how well they would perform on average ($Mean_{predict} = 66.26$, $Median_{predict} = 61.14$, $SD_{predict} = 28.63$; $Mean_{learn} = 87.28$, $Median_{learn} = 86.58$, $SD_{learn} = 21.60$; $V = 1293$, $p_{FDR\text{-}corrected} < .001$), at the start ($Mean_{predict} = 98.55$, $Median_{predict} = 88.93$, $SD_{predict} = 44.32$; $Mean_{learn} = 121.85$, $Median_{learn} = 109.42$, $SD_{learn} = 45.55$; $V = 2067$, $p_{FDR\text{-}corrected} < .001$), and at the end of the game ($Mean_{predict} = 51.37$, $Median_{predict} = 46.62$, $SD_{predict} = 29.58$; $Mean_{learn} = 74.89$, $Median_{learn} = 70.58$, $SD_{learn} = 25.35$; $V = 1430$, $p_{FDR\text{-}corrected} < .001$). They also underestimated the variance in their performance across 50 trials ($Mean_{predict} = 39.58$, $Median_{predict} = 35.46$, $SD_{predict} = 17.53$; $Mean_{learn} = 60.45$, $Median_{learn} = 58.38$, $SD_{learn} = 21.98$; $V = 769$, $p_{FDR\text{-}corrected} < .001$) and their learning rates, predicting that they would improve slower than they did ($Mean_{predict} = -0.17$, $Median_{predict} = -0.08$, $SD_{predict} = 0.23$; $Mean_{learn} = -0.28$, $Median_{learn} = -0.13$, $SD_{learn} = 0.32$; $V = 4773$, $p_{FDR\text{-}corrected} = .02$; learning rates extracted excluding the practice trials such that the predicted and actual learning trials matched at 50 trials each; Fig. 3c).

Different from results in Experiment 3a, participants' predicted average performance, starting performance, and variance tracked their actual performance (average performance: $rho = 0.24$, $p_{FDR\text{-}corrected} = .02$; starting performance: $rho = 0.20$, $p_{FDR\text{-}corrected} = .04$; variance: $rho = 0.25$, $p_{FDR\text{-}corrected} = .02$), showing some enhanced metacognitive awareness of performance with practice. However, as in Experiment 3a, no significant correlation was found between participants' predicted and actual final performance ($rho = 0.09$, $p_{FDR\text{-}corrected} = .4$) or learning rates ($rho = -0.05$, $p_{FDR\text{-}corrected} = .6$).

To explore whether the presence of three practice trials prior to making predictions improved prediction accuracy, we compared prediction errors in the average, starting, and final performance, as well as variance and learning rates in Experiments 3a and 3b. Using Wilcoxon rank-sum tests, we observed a reduction of prediction errors in average performance ($Mean_{exp\ 3a} = -32.82$, $Median_{exp\ 3a} = -29.97$, $SD_{exp\ 3a} = 33.34$; $Mean_{exp\ 3b} = -21.02$, $Median_{exp\ 3b} = -22.98$, $SD_{exp\ 3b} = 31.86$; $W = 6125$, $p_{FDR\text{-}corrected} = .02$), starting performance ($Mean_{exp\ 3a} = -67.85$, $Median_{exp\ 3a} = -57.29$, $SD_{exp\ 3a} = 79.58$; $Mean_{exp\ 3b} = -23.30$, $Median_{exp\ 3b} = -19.19$, $SD_{exp\ 3b} = 57.20$; $W = 4762$, $p_{FDR\text{-}corrected} < .001$), and variance ($Mean_{exp\ 3a} = -32.12$, $Median_{exp\ 3a} = -25.30$, $SD_{exp\ 3a} = 29.10$; $Mean_{exp\ 3b} = -20.86$, $Median_{exp\ 3b} = -17.01$, $SD_{exp\ 3b} = 24.81$; $W = 5843$, $p_{FDR\text{-}corrected} = .009$) with the addition of practice. We also found a trend for smaller prediction errors in final performance ($Mean_{exp\ 3a} = -33.61$, $Median_{exp\ 3a} = -31.75$, $SD_{exp\ 3a} = 35.39$; $Mean_{exp\ 3b} = -23.52$, $Median_{exp\ 3b} = -25.46$, $SD_{exp\ 3b} = 37.11$; $W = 6371$, $p_{FDR\text{-}corrected} = .052$), but no reduction in learning rates ($Mean_{exp\ 3a} = 0.15$, $Median_{exp\ 3a} = 0.12$, $SD_{exp\ 3a} = 0.41$; $Mean_{exp\ 3b} = 0.11$, $Median_{exp\ 3b} = 0.01$, $SD_{exp\ 3b} = 0.41$; $W = 8428$, $p_{FDR\text{-}corrected} = .2$) with the addition of practice.

Allowing participants to practice prior to making predictions also effectively eliminated the sharp decline in their affect ratings from baseline to the first round identified in Experiment 3a (Fig. 4a). Participants' affect rating changes between this interval did not differ significantly from 0 ($Mean = -5.2$, $Median = -1$, $SD = 24.64$, one-sample Wilcoxon $V = 3127.5$, $p = .1$) and were smaller compared to Experiment 3a ($Mean = -16.04$, $Median = -14$, $SD = 27.36$; Wilcoxon $W = 9804$, $p < .001$).

6.2.2. Exploratory analyses

We examined if participants' actual performance or prediction error in the first round would better predict changes in their own affect ratings. Consistent with Experiment 3a, prediction error was a better predictor of changes in affect and had a lower *AIC* score compared to actual performance in the first round (prediction error: $b = 0.17$, $p_{FDR\text{-}corrected} < .001$, $AIC = 774.66$; actual performance: $b = -0.16$, $p_{FDR\text{-}corrected} = .001$, $AIC = 786.58$; see full regression tables in SI; Fig. 4b and c).

6.3. Interim discussion

As in Experiment 3a, people's actual and predicted learning trajectories were better fit by an exponential decay function compared to a linear function. Critically, we found that just three practice trials made people's future performance predictions more accurate (although they were still inflated) and eliminated the sharp affect drop found at the beginning of learning *without* practice in Experiment 3a. Note that practice did not change participant's predictions of their future learning rate. In sum, brief practice improves the accuracy of people's predictions of their future performance which in turn may help participants remain positive at the beginning of skill acquisition.

7. General discussion

We found that people accurately predict the exponential decay shape of future skill learning curves on a novel visuomotor task prior to any task experience both for themselves and imagined naïve players. Critically, people do not *always* predict that learning curves will follow an exponential decay function in a bounded time frame; rather, people only predict that learning unfolds exponentially when a player puts in effort and the task is not too difficult. Although people correctly predict the shape of skill learning curves, they misrepresent the specific parameters, both for others and themselves — they are overly optimistic about future performance but underestimate learning rates. Our results suggest that these inaccuracies may be consequential for learning by fueling negative feelings at the beginning of skill acquisition, when predicted performance exceeds actual performance. Importantly, we found that practice can ameliorate this effect: After just three practice trials, people's performance predictions become more accurate and in turn, they feel better at the beginning of learning.

In contrast to prior work showing that people tend to linearize exponential functions (the Exponential Growth Bias, Wagenaar & Sagaria, 1975; McKenzie & Liersch, 2011; Stango & Zinman, 2009), our work revealed that people correctly intuit the exponential decay shape of learning curves. Participants predicted initial rapid improvement followed by a longer period of slower improvement, both when providing explicit trial-by-trial point estimates and more abstract likelihood estimates of future learning curves. The discrepancy between our findings and prior work on exponential thinking may be due to how we asked people to make predictions. For instance, in the classic duckweed and pond paradigm, Wagenaar and Timmers (1978) asked participants to indicate "the proportion of elapsed time" in comparison to when the pond will be fully filled. This task question is very cognitively demanding, requiring participants to reason and connect abstract properties like time and growth. In our paradigm, participants were asked to do something more intuitive — they directly simulated a series of performance data at specific moments in time. This is one key benefit of using a motor paradigm that has overt, explicit, and tangible trial-by-trial performance metrics, rather than a more abstract, cognitive paradigm (e.g., word list memory tasks). In contrast to prior work, we also asked participants to make predictions on relatively shorter time scales (e.g., performance over 50 trials vs. retirement savings over 40 years; McKenzie & Liersch, 2011), which recent work has shown makes exponential reasoning more accurate (e.g., how much will COVID-19 cases increase in 3 days vs. 15 days; Lammers, Crusius, & Gast, 2020). We also minimized linguistic cue-based reasoning, which can distort predictions of learning curves (Ariel et al., 2014).

Our work suggests that people do not simply apply a "learning is exponential" heuristic when thinking about the shape of future learning curves, but rather adjust their estimates based on their beliefs about how task difficulty and players' effort relate to performance gains. In line with prior theoretical and empirical work (Heller et al., 2015; Hodges & Lohse, 2020), people only predict that learning curves will follow an exponential decay function in a bounded time frame when the task is at an appropriate difficulty level and the learner puts in some effort. As such, people may be incorporating a priori beliefs about learning with observed features of the task, and the observed or imagined motivational state of the player, when making performance predictions. This approach is in line with "theory-based judgments" from the JOL literature whereby people use their general knowledge about learning to predict future performance (Nisbett & Wilson, 1977; Koriat et al., 2004; Rhodes, 2016; Mueller & Dunlosky, 2017). Of course, it is likely that cues beyond effort and difficulty, like a player's competence or external rewards (Howard, Bureau, Guay, Chong, & Ryan, 2021; Schneider, 1998), inform people's predictions. Future work is needed to explore how adults integrate multiple cues related to performance, both explicit and implicit, to construct learning curves.

Although people's predictions of their own skill learning curves in our task match the shape of their actual skill learning curves, they systematically misrepresent the precise parameters of their own and other people's learning curves. Specifically, individuals think they and others will start and end better than they actually do in our task and, in turn, that they and others will improve at a slower rate than in actuality. Many factors could lead to these discrepancies, including ones specific to our task. Adults' general optimism may inflate performance predictions, as prior work shows that adults can be overly positive about their future self on questionnaires as well as on tasks that probe beliefs and performance (Garrett & Sharot, 2017; Lefebvre et al, 2017; Sharot, Korn, & Dolan, 2011). Other biases, such as the "planning fallacy", might cause people to underestimate the time or effort needed to reach specific performance goals (Kahneman & Tversky, 1977). Players may have also misperceived the task difficulty (e.g., how hard it would be to learn how long to optimally hold the "enter" key to toss the lollipop), or their own skill, given that they never played the game before. Indeed, three practice trials prior to making performance predictions reduced prediction errors, in line with prior research showing that practice improves forecasting (Finn & Metcalfe, 2007; Finn & Metcalfe, 2008; Horn & Loewenstein, 2024; Koriat et al., 2002; Koriat & Bjork, 2005; Kornell & Metcalfe, 2006). Although practice makes people's performance predictions more accurate, these predictions are still inflated. Thus, people's systematic overestimation of performance and underestimation of learning rates across experiments are not simply a by-product of poorly calibrated beliefs about task difficulty or one's skill. Precisely how much experience is needed for accurate learning curve predictions is an area for future research.

Misperceptions of specific features of one's own learning curve may have significant consequences for actual learning outcomes. Initial optimism may be helpful by inspiring people to try new tasks. However, this initial optimism may be problematic once people embark on skill learning: We found that the more optimistic people were about their future learning progress, the more their mood dropped after the first round. This finding is in line with prior work showing that slowly improving on a task that you expect to quickly master can lead to quitting (see Dai et al., 2018; Lee & Wishart, 2005). Our task did not allow our participants to quit. If it did, their sudden drop in mood after the first five trials may have resulted in premature quitting, as their performance and their mood ultimately did improve. Across trials, participant mood and motivation were better fit by actual performance, rather than prediction errors between actual and predicted performance, potentially because people forgot their predictions over time or because their actual performance was more salient. Thus, it may be that predicted learning curves are most impactful at the very beginning of learning, when people are adjusting how to allocate their effort. Importantly, it is reasonable to assume that the very beginning of learning is actually the most critical period — this is likely when people make thin-slice judgments about their competency on a task, and perhaps decide to quit. Our work shows that low-stakes practice may serve as an effective intervention since it reduces people's prediction errors and negative affect changes at the beginning of learning.

Although we showed that adults can intuit learning curves prior to task engagement, it is unclear whether people spontaneously mentally construct future learning curves in their daily life without prompting. Going back to our opening example about learning a musical instrument, do people naturally think about their learning curves before starting a task or while engaging in one? Past work on cognitive monitoring has shown that people do spontaneously monitor task demands and their own performance and opt to engage in easier tasks (Kool & Botvinick, 2018; Niebaum, Chevalier, Guild, & Munakata, 2019; Shenhav et al., 2017). Moreover, learners adaptively choose to study items of desirable difficulty and choose to take on more difficult tasks only when they anticipate swift improvement (Bjork & Bjork, 2020; Moskowitz et al., 2020). Our work suggests that people may think about their expected progress before starting a new task and choose tasks according to what they think is tractable. At the same time, mentally simulating entire learning curves for new tasks may itself be cognitively demanding. Thus, learners might not engage in elaborated learning curve prediction spontaneously; it is possible that learners might only predict the immediate performance gain (such as "I will play through the first five chords without errors next time"), hold a more general representation of progress ("I will get better with practice"), or only represent the end goal ("I will master playing this song"). Considering that making predictions about future outcomes may both improve learning (Brod, Hasselhorn, & Bunge, 2018) and goal-setting, future work should explore how learners with different experience levels may intuit their future progress in more naturalistic settings, as well as if prompting learners to think about their learning progress in a more fine-grained manner improves learning outcomes.

Our findings also speak to motivational and pedagogical theories. First, by showing that people expect that practice will lead to swift improvements in ability when someone puts in effort and the task is not too hard, our work highlights the circumstances under which people might possess a "growth mindset" (Dweck, 2006; Jia, Lim, Ismail, & Tan, 2021; Walton & Yeager, 2020; Yeager et al., 2019). Furthermore, by assessing predictions of performance on a trial-by-trial basis, we revealed with significant granularity people's perceptions about the relationship between practice and ability in the context of skill learning. Moving forward, our approach could be used to gain a better understanding of people's task-specific mindsets — if they have a fixed mindset about learning to sing, would they predict a flatter learning curve? Second, our research also points to where teaching and learning can go awry. Teachers constantly think about which tasks are most appropriate for learners, where students should be along their learning curve, and how to provide better support for learning (Bridgers, Jara-Ettinger, & Gweon, 2020; Gweon, 2021; Popp & Gureckis, 2020; Shafto, Goodman, & Griffiths, 2014). Our work suggests that if teachers misestimate a student's effort or a task's difficulty, they risk mis-calibrating their estimate of the learner's true learning curve, and in turn, provide sub-optimal pedagogy. Furthermore, we show that learners themselves miscalibrate their starting point and rate of learning, both of which may necessitate intervention from a teacher to better scaffold expectations.

The question of how people think about learning curves may arguably be most important to study in children, who are rapidly acquiring foundational knowledge. Recent work shows that even 4- to 8-year-old children predict gradual improvement in the first few trials on a novel skill learning task (Zhang et al., n.d.). However, it remains unknown when in development children hold adult-like intuitions about the shape of learning curves on a longer timescale (e.g., that performance quickly improves early on and then slows down in skill learning; the exponential decay shape). Given that children are even more optimistic about their future performance than adults (Schneider, 1998; Lockhart, Goddu, & Keil, 2021; Leonard & Sommerville, 2025; Xia, Poorthuis, & Thomaes, 2024), children may over-predict their starting and average performance and under-predict their learning rates more than adults. Critically, our work suggests that children's over optimism could have negative consequences: Since large prediction errors at the beginning of learning relate to negative affect, children may be even more likely than adults to feel bad and quit at the beginning of learning. Future work is necessary to explore (1) when children have the cognitive capacity to predict their future learning curves and (2) how these predictions relate to motivation and learning.

Our study has a number of limitations. First, we only tested predictions of learning curves on one novel visuomotor task. It is unknown whether people predict that learning curves will unfold following an exponential decay function on other motor tasks, or on more abstract, cognitive tasks. Answering this question for more opaque cognitive tasks may be challenging as work on learning trajectories in the cognitive domain has been relatively sparse and more heterogeneous (Tulving, 1967; Rice, Wexler, & Hershberger, 1998; Son & Sethi, 2010; Howard, 2018). Thus, further research is needed to examine whether people can predict the varied shapes of learning curves across domains and on different time scales. Second, our task imposed a bounded time scale of 50 trials for the learning curve predictions, which only captured the beginning process of learning on this task. It is an open question, for example, whether people might predict exponential decay shapes of learning trajectories for a player who puts in effort on a *difficult* task in Experiment 2 over a longer time scale. Third, although we explored people's representations of learning curves, we never explicitly labeled them as such. It is unclear how people explicitly think about "learning curves" per se. Fourth, our use of a within-subjects design in Experiment 2, allowed us to reduce participant-level variability and increase power, but this design may have also inadvertently increased the likelihood that participants suspected what the goal of the experiment was. Thus, future work should run a between-subjects version of Experiment 2 to fully rule out demand effects. This work would be further bolstered by having participants actually play the Lolli-toss tasks with varying difficulty and varying effort to confirm that actual learning curves differ by these conditions. Finally, we recruited participants from a convenience sample in a Western, educated, industrialized, rich, and democratic (WEIRD) country, so we cannot address whether our findings generalize across socio-cultural contexts.

We talk and think a lot about learning curves. Coaches and educators give advice about learning curves (tennis coach Brad Gilbert has said: "There's always a learning curve, where you've got to learn what your subject is all about"; Drucker, 2007), pop-psychology books are written about them (e.g., the so-called "ten thousand hour rule" in Outliers by Malcom Gladwell, 2008), and they are even mentioned in children's literature ("But I do think that when people say 'a learning curve,' they make a mistake. Learning to me always seems to go in a straight, ignorant line and then, every so often, takes a jump straight upward." Jones, 2011). Yet, it is not always clear what people mean when they talk about learning curves or whether they have an accurate sense of them. Here we show that, on a novel skill learning task, people accurately predict that learning curves typically start off swiftly and then level off — a prediction made *before* they even engaged with the task. This apparent intuitive understanding of learning curves may have a direct relationship to how people make crucial decisions about what tasks to pursue in the first place, and when to persist versus quit. A better understanding of people's mental representations of learning curves thus has wide implications for studies of development, pedagogy, and real-life decision-making.

CRediT authorship contribution statement

Xiuyuan Zhang: Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing, Writing – original draft. **Samuel D. McDougle:** Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization, Writing – review & editing. **Julia A. Leonard:** Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization, Writing – review & editing, Writing – original draft.

Declaration of competing interest

All authors declare no conflict of interest.

Data availability

All experiments are preregistered. Materials, preregistrations, data, and analysis code for this manuscript can be accessed on OSF repository: https://osf.io/xzm5c.

Acknowledgement

We thank Mika Asaba, Frank Keil, Joshua Knobe, Darko Odic, Aalap Shah, Elaine Wang, Kimberly Wong, and Sami Yousif for their feedback and discussions, and research assistance from Bethel Asomaning, Saif Behairy, Adriana Christakis, Lizbeth Lozano, and Christina Norberg. We also thank members of the Yale Leonard Learning Lab and Yale Cognition and Development Lab for helpful discussions and feedback. This research was supported by a Jacobs Foundation Research Fellowship to J.A.L.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106083.

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*(6631), 401–406. https://doi.org/10.1038/387401a0

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Anderson, D. I., Lohse, K. R., Lopes, T. C. V., & Williams, A. M. (2021). Individual differences in motor skill learning: Past, present and future. *Human Movement Science, 78*, Article 102818. https://doi.org/10.1016/j.humov.2021.102818

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*(1), 126. https://doi.org/10.1037/h0027455

Ariel, R., Hines, J. C., & Hertzog, C. (2014). Test framing generates a stability bias for predictions of learning by causing people to discount their learning beliefs. *Journal of Memory and Language, 75*, 181–198. https://doi.org/10.1016/j.jml.2014.06.003
Berke, M., Tenenbaum, A., Sterling, B., & Jara-Ettinger, J. (2023). Thinking about thinking as rational computation. In *Proceedings of the 45th annual meeting of the cognitive science society*. https://escholarship.org/uc/item/93b7j877.
Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition, 9*(4), 475. https://doi.org/10.1016/j.jarmac.2020.09.003
Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology, 64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823
Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2020). Young children consider the expected utility of others' learning to decide what to teach. *Nature Human Behaviour, 4*(2), 144–152. https://doi.org/10.1038/s41562-019-0748-6
Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction, 55*, 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013
Ciccione, L., Sablé-Meyer, M., & Dehaene, S. (2022). Analyzing the misperception of exponential growth in graphs. *Cognition, 225*, Article 105112. https://doi.org/10.1016/j.cognition.2022.105112
Dai, H., Dietvorst, B. J., Tuckfield, B., Milkman, K. L., & Schweitzer, M. E. (2018). Quitting when the going gets tough: A downside of high performance expectations. *Academy of Management Journal, 61*(5), 1667–1691. https://doi.org/10.5465/amj.2014.1045
Drucker, J. (2007, February 14). *Two coaches, two styles, one goal*. ESPN. Retrieved July 12, 2024, from https://www.espn.com/sports/tennis/news/story?id=2765737.
Duckworth, A. L., Eichstaedt, J. C., & Ungar, L. H. (2015). The mechanics of human achievement. *Social and Personality Psychology Compass, 9*(7), 359–369. https://doi.org/10.1111/spc3.12178
Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random house.
Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 238. https://doi.org/10.1037/0278-7393.33.1.238
Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58*(1), 19–34. https://doi.org/10.1016/j.jml.2007.03.006
Frömer, R., Stürmer, B., & Sommer, W. (2016). (Don't) mind the effort: Effects of contextual interference on ERP indicators of motor preparation. *Psychophysiology, 53*(10), 1577–1586. https://doi.org/10.1111/psyp.12703
Garrett, N., & Sharot, T. (2017). Optimistic update bias holds firm: Three tests of robustness following Shah et al. *Consciousness and Cognition, 50*, 12–22. https://doi.org/10.1016/j.concog.2016.10.013
Gladwell, M. (2008). *Outliers: The story of success*. Little, Brown and Co.
Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior, 36*(2), 212–224. https://doi.org/10.3200/JMBR.36.2.212-224
Gweon, H. (2021). Inferential social learning: Cognitive foundations of human social learning and teaching. *Trends in Cognitive Sciences, 25*(10), 896–910. https://doi.org/10.1016/j.tics.2021.07.008
Gweon, H., Asaba, M., & Bennett-Pierre, G. (2017). Reverse-engineering the process: Adults and preschoolers' ability to infer the difficulty of novel tasks. In *Proceedings of the 39th annual conference of the cognitive science society*.
Heathcote, A., Brown, S., & Mewhort, D. J. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*(2), 185–207. https://doi.org/10.3758/BF03212979
Heller, D., Arnold, J. E., Klein, N., & Tanenhaus, M. K. (2015). Inferring difficulty: Flexibility in the real-time processing of disfluency. *Language and Speech, 58*(2), 190–203. https://doi.org/10.1177/0023830914528107
Hodges, N. J., & Lohse, K. R. (2020). Difficulty is a real challenge: A perspective on the role of cognitive effort in motor skill learning. *Journal of Applied Research in Memory and Cognition, 9*(4), 455. https://doi.org/10.1016/j.jarmac.2020.08.006
Horn, S., & Loewenstein, G. (2024). *Underestimating learning by doing*. Available at SSRN: https://doi.org/10.2139/ssrn.3941441.
Howard, J. L., Bureau, J. S., Guay, F., Chong, J. X., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science, 16*(6), 1300–1323. https://doi.org/10.1177/1745691620966789
Howard, R. W. (2018). Development of chess skill from domain entry to near asymptote. *The American Journal of Psychology, 131*(3), 323–345. https://doi.org/10.5406/amerjpsyc.131.3.0323
Hutzler, F., Richlan, F., Leitner, M. C., Schuster, S., Braun, M., & Hawelka, S. (2021). Anticipating trajectories of exponential growth. *Royal Society Open Science, 8*(4), Article 201574. https://doi.org/10.1098/rsos.201574
Jia, L., Lim, C. H., Ismail, I., & Tan, Y. C. (2021). Stunted upward mobility in a learning environment reduces the academic benefits of growth mindsets. *Proceedings of the National Academy of Sciences, 118*(10), Article e2011832118. https://doi.org/10.1073/pnas.2011832118
Jones, D. W. (2011). *Enchanted glass*. Greenwillow Books.
Kahneman, D., & Tversky, A. (1977). Intuituve prediction. In *Biases and corrective procedures* (p. 0044). Decision Research, Perceptronics.
Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour, 2*(12), 899–908. https://doi.org/10.1038/s41562-018-0401-9
Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349. https://doi.org/10.1037/0096-3445.126.4.349
Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187. https://doi.org/10.1037/0278-7393.31.2.187
Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1133. https://doi.org/10.1037/0278-7393.32.5.1133
Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting One's Own Forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643
Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107. https://doi.org/10.1037/0278-7393.6.2.107
Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147–162. https://doi.org/10.1037/0096-3445.131.2.147
Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: overestimating remembering and underestimating learning. *Journal of experimental psychology: General, 138*(4), 449. https://doi.org/10.1037/a0017350
Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 609. https://doi.org/10.1037/0278-7393.32.3.609
Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological science, 22*(6), 787–794. https://doi.org/10.1177/0956797611407929
Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., & Haith, A. M. (2019). Motor learning. *Comprehensive Physiology, 9*(2), 613–663. https://doi.org/10.1002/cphy.c170043
Lammers, J., Crusius, J., & Gast, A. (2020). Correcting misperceptions of exponential coronavirus growth increases support for social distancing. *Proceedings of the National Academy of Sciences, 117*(28), 16264–16266. https://doi.org/10.1073/pnas.2006048117
Lee, T. D., Swinnen, S. P., & Serrien, D. J. (1994). Cognitive effort and motor learning. *Quest, 46*(3), 328–344. https://doi.org/10.1080/00336297.1994.10484130
Lee, T. D., & Wishart, L. R. (2005). Motor learning conundrums (and possible solutions). *Quest, 57*(1), 67–78. https://doi.org/10.1080/00336297.2005.10491843
Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour, 1*(4), 0067. https://doi.org/10.1038/s41562-017-0067
Leonard, J. A., & Sommerville, J. A. (2025). A unified account of why optimism declines in childhood. *Nature Reviews Psychology, 4*, 35–48. https://doi.org/10.1038/s44159-024-00384-z
Lockhart, K. L., Goddu, M. K., & Keil, F. C. (2021). How much can you learn in one year? how content, pedagogical resources, and learner's age influence beliefs about knowledge acquisition. *Cognitive Development, 60*, 101115. https://doi.org/10.1016/j.cogdev.2021.101115
McKenzie, C. R., & Liersch, M. J. (2011). Misunderstanding savings growth: Implications for retirement savings behavior. *Journal of Marketing Research, 48*(SPL), S1–S13. https://doi.org/10.1509/jmkr.48.SPL.S1
Melnik-Leroy, G. A., Aidokas, L., Dzemyda, G., Dzemydaitė, G., Marcinkevičius, V., Tiešis, V., & Usovaitė, A. (2023). Is my visualization better than yours? Analyzing factors modulating exponential growth bias in graphs. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1125810, 1125810.
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
Moskowitz, J. B., Gale, D. J., Gallivan, J. P., Wolpert, D. M., & Flanagan, J. R. (2020). Human decision making anticipates future performance in motor learning. *PLOS Computational Biology, 16*(2), Article e1007632. https://doi.org/10.1371/journal.pcbi.1007632
Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93*, 245–258. https://doi.org/10.1016/j.jml.2016.10.008
Muradoglu, M., & Cimpian, A. (2020). Children's intuitive theories of academic performance. *Child Development, 91*(4), e902–e918. https://doi.org/10.1111/cdev.13325
Newell, A., & Rosenbloom, P. S. (2013). Mechanisms of skill acquisition and the law of practice. In *Cognitive skills and their acquisition* (pp. 1–55). Psychology Press.
Niebaum, J. C., Chevalier, N., Guild, R. M., & Munakata, Y. (2019). Adaptive control and the avoidance of cognitive control demands across development. *Neuropsychologia, 123*, 152–158. https://doi.org/10.1016/j.neuropsychologia.2018.04.029
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231. https://doi.org/10.1037/0033-295X.84.3.231
Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General, 143*(1), 255. https://doi.org/10.1037/a0030825
Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*(3), 105–119. https://doi.org/10.1111/j.1539-6053.2009.01038.x
Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology, 92*(1), 128. https://doi.org/10.1037/0021-9010.92.1.128

Popp, P. O., & Gureckis, T. M. (2020). Ask or Tell: Balancing questions and instructions in intuitive teaching. In *Proceedings of the 42nd annual conference of the cognitive science society*.

Rhodes, M. G. (2016). *Judgments of learning: Methods, data, and theory. The Oxford handbook of metamemory, 1*, 65–80. https://doi.org/10.1093/oxfordhb/9780199336746.013.4

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1412–1431. https://doi.org/10.1044/jslhr.4106.1412

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences, 111*(33), 12252–12257. https://doi.org/10.1073/pnas.1407535111

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

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology, 71*, 55–89. https://doi.org/10.1016/j.cogpsych.2013.12.004

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience, 14*(11), 1475–1479. https://doi.org/10.1038/nn.2949

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience, 40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(4), 907. https://doi.org/10.1037/0278-7393.27.4.907

Solum, M., Lorås, H., & Pedersen, A. V. (2020). A golden age for motor skill learning? Learning of an unfamiliar motor task in 10-year-olds, young adults, and adults, when starting from similar baselines. *Frontiers in Psychology, 11*, 538. https://doi.org/10.3389/fpsyg.2020.00538

Son, L. K., & Sethi, R. (2010). Adaptive learning and the allocation of time. *Adaptive Behavior, 18*(2), 132–140. https://doi.org/10.1177/1059712309344776

Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science, 19*(10), 986–988. https://doi.org/10.1111/j.1467-9280.2008.02189.x

Stango, V., & Zinman, J. (2009). Exponential growth bias and household finance. *The Journal of Finance, 64*(6), 2807–2849. https://doi.org/10.1111/j.1540-6261.2009.01518.x

Ten, A., Kaushik, P., Oudeyer, P. Y., & Gottlieb, J. (2021). Humans monitor learning progress in curiosity-driven exploration. *Nature Communications, 12*(1), 5972. https://doi.org/10.1038/s41467-021-26196-w

Thorndike, E. L. (1913). *The psychology of learning* (Vol. 2). Teachers College: Columbia University.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of verbal learning and verbal behavior, 6*(2), 175–184. https://doi.org/10.1016/S0022-5371(67)80092-6

Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential growth. *Perception & Psychophysics, 18*, 416–422. https://doi.org/10.3758/BF03204114

Wagenaar, W. A., & Timmers, H. (1978). Extrapolation of exponential time series is not enhanced by having more data points. *Perception & Psychophysics, 24*, 182–184. https://doi.org/10.3758/BF03199548

Walton, G. M., & Yeager, D. S. (2020). Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science, 29*(3), 219–226. https://doi.org/10.1177/0963721420904453

Xia, M., Poorthuis, A. M. G., & Thomaes, S. (2024). Children's overestimation of performance across age, task, and historical time: A meta-analysis. *Child Development, 95*, 1001–1022. https://doi.org/10.1111/cdev.14042

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature, 573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y

Yildirim, I., Saeed, B., Bennett-Pierre, G., Gerstenberg, T., Tenenbaum, J., & Gweon, H. *Explaining intuitive difficulty judgments by modeling physical effort and risk.*. (2019). arXiv preprint arXiv:1905.04445.

Zhang, X., Carrillo, B.A., Christakis, A., Leonard. J.A. (in press) Children predict improvement on novel skill learning tasks. *Child Development*.