Cognition 266 (2026) 106340

Contents lists available at ScienceDirect

Cognition

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

Full Length Article

Children leverage predictive representations for flexible, value-guided choice

Alice Zhang [a], Ari E. Kahn [b], Nathaniel D. Daw [b,c], Kate Nussenbaum [a,b,1], Catherine A. Hartley [a,*,1]

[a] Department of Psychology, New York University, New York, NY, USA
[b] Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA
[c] Department of Psychology, Princeton University, Princeton, NJ, USA

ARTICLE INFO

Keywords:
Cognitive development
Value-guided choice
Successor representation
Model-based reinforcement learning
Reward revaluation

ABSTRACT

By harnessing a mental model of how the world works, learners can make flexible choices in changing environments. However, while children and adolescents readily acquire structured knowledge of their environments, relative to adults, they often demonstrate weaker signatures of leveraging this knowledge to plan actions. One explanation for these developmental differences is that using a mental model to prospectively simulate potential choices and their outcomes is computationally costly, taxing cognitive mechanisms that develop into adulthood. Here, we ask whether children effectively leverage structured knowledge to make flexible choices by relying on two alternative strategies that do not require costly mental simulation at choice time. First, through offline replanning, models can be queried before the time of choice to update the values of potential actions. Second, an abstracted predictive model, known as a successor representation (SR), can enable simplified computation of long-run reward values of candidate actions without requiring iterative simulation of multiple time steps. Here, across three experiments, we assessed whether children, adolescents, and adults aged 7–23 years similarly harness these learning strategies. In a reward revaluation task, we found that children flexibly updated their behavior by leveraging structured knowledge, but that across age, the opportunity for offline replanning during rest did not influence behavior. While participants may have leveraged a detailed mental model of the task structure, they may have also relied on simplified, predictive representations to guide their choices. We then directly tested whether children use predictive representations and observed early-emerging use of the SR, providing a mechanistic account of how children use structured knowledge to guide choice without detailed model-based simulation.

1. Introduction

To make good choices in a richly structured and changing world, people can learn and exploit relations between different actions and events to guide their decisions. By relying on knowledge of the environment (an "internal model") to mentally simulate different sequences of actions and the outcomes they are likely to yield (Balleine & O'Doherty, 2010; Daw et al., 2005; Doll et al., 2015; Vikbladh et al., 2024), learners can flexibly update their beliefs in the absence of direct experience. This form of learning, known as "model-based" learning, enables rapid adaptation to changing environments, though with a high computational cost. Learners can also update the estimated values of their actions experientially, based on the outcomes that their actions ultimately yield. However, while this "model-free" form of learning is computationally efficient, it is also rigid — if the environment changes, a model-free learner can only update their estimated values for different actions by taking those actions and experiencing their new consequences. Evidence suggests that human learners exploit both model-based and model-free approaches, trading off flexibility and efficiency across different environments based on the demands of the learning problems they face (Daw et al., 2005; Kool et al., 2017). Recent work further suggests that beyond simply switching between these two forms of learning, adults also exploit alternative learning and decision strategies that balance the efficiency of model-free learning with the flexibility

* Corresponding author.
E-mail address: cate@nyu.edu (C.A. Hartley).
[1] Shared senior authorship.

https://doi.org/10.1016/j.cognition.2025.106340
Received 28 February 2025; Received in revised form 25 August 2025; Accepted 4 October 2025
Available online 10 October 2025

of model-based learning (Collins & Cockburn, 2020; Dolan & Dayan, 2013; Doll et al., 2012; Keramati et al., 2016). To date, however, it is unclear whether children and adolescents also use these "intermediate" learning strategies to guide their choices.

Like adults, children and adolescents frequently experience changing environments in which flexible decision making can be facilitated through the use of structured knowledge. However, the strategies that people use to make rewarding choices change across development (Raab & Hartley, 2018). While signatures of model-free learning strategies remain relatively consistent across age, evidence of model-based learning increases into adolescence and early adulthood (Cohen et al., 2020; Decker et al., 2016; Nussenbaum, Scheuplein, et al., 2020; Palminteri et al., 2016; Potter et al., 2017; Smid, Ganesan, et al., 2023). However, even in contexts in which children fail to use world models to guide their choices, in many cases, they still acquire structured knowledge about their learning environments that is revealed in other ways. For instance, they can explicitly report the states to which their actions may lead (Decker et al., 2016; Nussenbaum, Scheuplein, et al., 2020; Potter et al., 2017) or the likely causal source of good and bad outcomes (Cohen et al., 2020).

This developmental dissociation between learning a model of the world and using it to guide decision making is a puzzling phenomenon (Hartley et al., 2021) — why do children often fail to use the knowledge they have acquired? One possibility is that the iterative, forward simulation processes on which model-based decision making depends are computationally costly. By this, we mean that they require the engagement of proactive cognitive control and working memory — cognitive abilities that continue to develop into adolescence and early adulthood and facilitate greater capacity for manipulating information in mind and faster processing speeds (Amso et al., 2014; Luna, 2009). Indeed, previous work has observed age-related change in behavior consistent with increases in planning depth (Ma et al., 2022), suggesting that iterative, step-by-step simulation of sequences of actions and outcomes improves *and* increasingly facilitates people's choices from childhood to early adulthood. Previous work on causal reasoning has also shown that children tend to perform worse than adults on tasks that require mental simulation of counterfactual possibilities (Kominsky et al., 2021; Nussenbaum, Cohen, et al., 2020; Rafetseder et al., 2013). In addition, young children demonstrate signatures of model-based behavior in simpler tasks with little to no planning depth (Kenward et al., 2009; Klossek et al., 2008), where fewer sequential outcomes and actions would need to be mentally simulated. Taken together, this work suggests that children may learn complex models of the environment but not use them to the same extent or in the same way as adults, and furthermore that this difference is likely due to the cognitive demands involved in iteratively computing the consequences of different actions across multiple future timesteps. It may be that these limitations are, in part, due to limited decision time — in many prior studies of model-based learning (Decker et al., 2016; Nussenbaum, Scheuplein, et al., 2020), participants had only a short temporal window to make their choices, which may have been particularly consequential for younger participants with slower processing speeds.

If children's failure to use knowledge of their environments to the same extent as adults is due to limitations in mentally simulating decision trajectories, then they may be able to leverage structured knowledge to make rewarding decisions when the demand for rapid, online, step-by-step simulation is attenuated. In the present series of studies, we ask whether children use structured knowledge to flexibly guide their choices when learning in environments that do not require, at the time of choice, the costly mental simulation associated with "pure" model-based learning strategies. We consider two alternative strategies that have been shown to support decision making in adults: leveraging models for evaluation offline (before a choice is faced), and using abstracted models, such as the successor representation (SR), that collapse multiple timesteps. These two strategies may respectively help children overcome the time and capacity costs of iterative mental simulation.

While model-based learning algorithms posit that iterative, mental simulation occurs at the time of choice, mental models can also be used to simulate potential sequences of experiences and update value representations before the need to select an action (Sutton, 1991). We use the term *offline evaluation* to describe such mental simulation that occurs after reward receipt or during rest rather than at the moment of choice. Offline evaluation may enable the brain to exploit periods of rest for learning, reducing the need for online planning at the time of choice. Offline evaluation may involve "replay" or the reactivation of memories of previous experiences, enabling them to be linked to newly experienced reward outcomes (Lengyel & Dayan, 2007; Sutton, 1991). Extensive work using multi-step decision tasks has found that adults leverage mental models to flexibly update their choice behavior after rewards in the environment change (Boddez et al., 2011; Dickinson & Burke, 1996; Liljeholm & Balleine, 2009). It was initially assumed that forward planning at the time of choice accounted for this success, but it may be the case that adults "solve" these tasks by learning during "offline" periods of rest. Recent work that has manipulated the opportunity for offline evaluation (Gershman et al., 2014) and measured neural activity during rest periods within the task (Momennejad et al., 2018) suggests that adults do indeed leverage offline evaluation to support value-guided decision making.

In addition to offline evaluation, use of abstracted world models such as the successor representation (SR) also can enable behavioral flexibility without requiring iterative, online forward simulation (Dayan, 1993). What makes traditional model-based evaluation costly is the requirement to iteratively search through multiple steps to piece together the likely outcomes of candidate actions. The SR is a predictive representation that stores, for each state, aggregated (rather than individual step-by-step) expectations about the future states that will likely follow it at some later point, potentially after multiple steps. The SR can be used to guide choice by combining these future-state expectations with information about the value of each state. Critically, the SR's aggregation simplifies this process by removing the need for iterative, step-by-step simulation of the potential sequences of states that may be experienced following an initial choice. Prior work using a number of different tasks has demonstrated evidence that people both form such temporally abstracted representations, and use them to guide choices (Garvert et al., 2017; Gershman, 2018; Kahn & Daw, 2025; Momennejad, 2020; Momennejad et al., 2017; Russek et al., 2021). In reward revaluation tasks, for example, adults may update their behavior by relying on statistical knowledge about the final states they tend to experience after each initial choice, leveraging these predictive representations when rewards change to compute new action values without costly forward simulation of all intermediate time steps (Momennejad et al., 2017).

Evidence that children "fail" to use structured knowledge to guide their decisions to the same extent as adults comes largely from tasks that may not afford the use of intermediate learning strategies like offline evaluation or the use of abstracted representations. Adults' use of such strategies raises the intriguing possibility that children and adolescents may similarly be able to harness internal models to make good choices in environments that permit them to be used via less costly computations. In the present series of studies, we ask whether and how children and adolescents harness structured knowledge to guide their choices when they can rely on intermediate learning strategies that combine the efficiency of model-free learning with the flexibility of model-based computation. In Experiment 1, we demonstrate that children and adolescents flexibly update their behavior in a reward revaluation task, when given the opportunity for offline replay. This provides evidence that children can indeed learn and use structured task knowledge in certain contexts, and accords with prior work where model-based behavior was demonstrated at younger ages (Kenward et al., 2009; Klossek et al., 2008). In Experiment 2, we further probe whether offline replay during rest facilitates flexible replanning by removing the task's

rest period. Here, we found that the removal of the rest phase did not significantly affect participants' behavioral flexibility, suggesting that a different strategy, such as online model-based planning or the use of predictive representations, may underlie children's ability to leverage structured knowledge in this task. In our final experiment, we directly ask whether children and adolescents make use of predictive representations like the SR. Using a multi-trial reinforcement learning task to concurrently characterize the use of model-free, model-based, and SR-based strategies, we find evidence for use of SR-based strategies in children as young as 8 years old. This suggests that young children are able to learn predictive representations and use them to flexibly guide behavior in environments with changing rewards.

Together, our results across three studies demonstrate that children effectively leverage structured knowledge to guide decision making and that they can do so by relying on predictive representations like the successor representation. Results from these experiments help to resolve the puzzle of why children often demonstrate adult-like learning but reduced *use* of structured knowledge, and suggest that children do harness sophisticated, predictive representations to guide choice when the learning environments they face allow them to do so.

2. Experiments 1 and 2

2.1. Methods

2.1.1. Participants

2.1.1.1. Experiment 1. 119 participants between the ages of 7–23 years completed the experiment online, remotely and asynchronously, and were included in the analyses. An additional 37 participants completed the study but were excluded from all analyses for predefined exclusion criteria including interacting with their browser window more than 20 times during the study session ($n = 2$ Adolescents), failing to respond on more than 15 % of the 100 learning trials or more than 15 % of the 84 memory trials (n = 2 Adults), making four or more errors on the task comprehension questions ($n = 1$ Adolescent), failing more than five out of 16 attention-check trials ($n = 10$ Children, n = 2 Adolescents, n = 1 Adult; see task details below), failing to learn to criterion in the learning phase of the task (n = 10 Children, $n = 3$ Adolescents, $n = 5$ Adults; see task details below), or potential parental interference (n = 1 Child). Sample size was determined a priori based on past work on decision making in age-continuous developmental samples (Jones et al., 2014; Ma et al., 2022; Nussenbaum et al., 2023; Nussenbaum, Scheuplein, et al., 2020; Somerville et al., 2017). Of the 119 participants included in the analyses, 42 were children (7.02–12.99 years; Mean age = 10.14; n = 19 females), 35 were adolescents (13.03–17.82 years; Mean age = 15.47; n = 17 females), and 42 were adults (18.42–23.82 years; Mean age = 20.96; n = 22 females). The study took approximately 40 min to complete and participants were paid via a $10 Amazon gift card along with an additional bonus that ranged from $0–5 based on task performance.

Participants were recruited primarily via Facebook and Instagram ads, as well as via word-of-mouth, local events, and flyers distributed around NYU. Prior to being eligible to participate in the online study, all participants were pre-screened in a 5-min video call with a researcher, during which they were required to be on camera and to confirm their full name and date of birth. Adult participants and parents of child and adolescent participants were additionally required to show photo identification. According to self- or parental-report, participants had normal or corrected-to-normal vision and no diagnosed psychiatric, neurodevelopmental, or learning disorders. 52.5 % of participants were White, 30.0 % were Asian, 8.3 % were Black, and 9.2 % were two or more races. In addition, 12.5 % of participants were Hispanic.

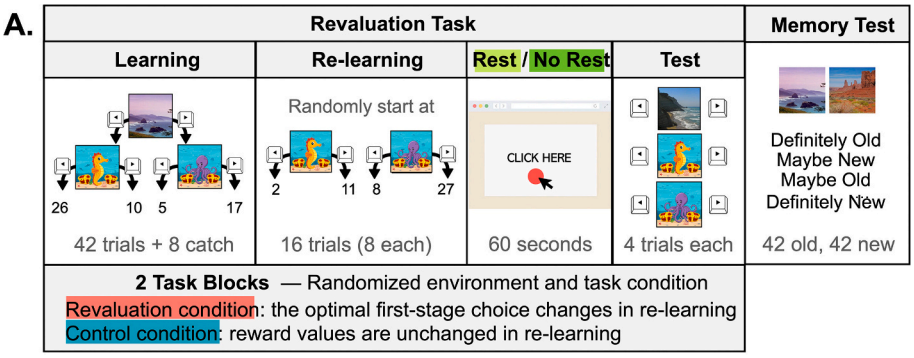
2.1.1.2. Experiment 2. 119 new participants between the ages of 7–23 years completed the task and were included in the analyses. Recruitment methods, exclusion criteria, and payment methods were the same as for Experiment 1, and participants who had already completed Experiment 1 were excluded from participating. Of the 119 participants included in the analyses, 41 participants were children (7.02–12.95 years; Mean age = 10.00, n = 18 females), 33 were adolescents (13.28–17.92 years; Mean age = 15.33, n = 15 females), and 45 were adults (18.05–24.00 years; Mean age = 20.90, n = 24 females). An additional 65 participants completed the study but were excluded for interacting with their browser window more than 20 times during the study session (n = 1 Child, n = 1 Adolescent, n = 2 Adults), failing to respond on more than 15 % of 100 learning trials or failing to respond on more than 15 % of 84 memory trials ($n = 4$ Children, n = 2 Adolescents, n = 1 Adult), making four or more errors on the task comprehension questions (n = 1 Children), failing more than five out of 16 attention-check trials ($n = 14$ Children, $n = 8$ Adolescents, n = 2 Adults), failing to learn to criterion in the learning phase of the task ($n = 20$ Children, $n = 5$ Adolescents, $n = 3$ Adults), or potential parental interference (n = 1 Children). According to self- or parental-report, 45 % of participants were White, 32.5 % were Asian, 7.5 % were Black, and 15 % were two or more races. In addition, 13.3 % of participants were Hispanic.

2.1.2. Experimental procedure

In both Experiments 1 and 2, participants completed an adapted version of a two-stage reward revaluation task used in prior adult work (Momennejad et al., 2018). The task was designed to assess whether participants across age relied on offline processing to flexibly update their choices when rewards in the environment changed. In the first stage of the task, participants learned to make two sequential choices to gain reward. Next, we elicited the need to update first-stage choices by changing the rewards in the environment. Participants experienced new rewards only in a "relearning" phrase of the task, where they did not make first-stage choices. In the final "test" phase, participants once again made first-stage choices. Here, we assessed whether participants chose the same first-stage choices that led to reward in the original learning phase, or whether they successfully "replanned" and used their knowledge of the new, second-stage rewards to update their first-stage choices. Critically, because participants did not make first-stage choices during relearning, replanning required leveraging their knowledge of the transition structure of the task. To gain insight into whether participants relied on offline processing to replan, we manipulated the opportunity for offline processing by including a rest phase after relearning in Experiment 1 but not Experiment 2.

The online task was programmed using jsPsych (de Leeuw, 2015) and hosted on Pavlovia. Our child-friendly task version was framed within a "Treasure Hunt" narrative, in which participants' overall goal was to find the most treasure. Participants completed two blocks of the task (Fig. 1A). In each task block, participants were told that they were in a specific environment (ocean or canyon) in which there were two different animals (seahorse and octopus, or lion and giraffe, respectively). Each animal had two treasure chests that contained different amounts of treasure (Fig. 1B).

As described above, the task consisted of four phases: learning, relearning, rest, and test. In the learning phase of the task, participants made 42 two-stage decisions to try to find the most treasure. Participants were told that their bonus payment would be contingent on how much treasure they found. On each trial, in the first decision stage, they saw a trial-unique image of the environment and had to choose to go either up or down to find an animal. Animals remained in the same position for the duration of the block. After reaching an animal, participants made a second-stage decision between the animal's left and right treasure chests. Each of the four chests had a different amount of treasure (between 5 and 50 pieces) that remained constant throughout the learning phase of each task block, allowing participants to learn to navigate to the most rewarding chest across trials. Participants had a time limit of 2 s to make each choice; if they did not make a choice within the allotted time,



B.

Animal	Chest	Original Value	Relearning Value (Control)	Relearning Value (Revaluation)
Animal 1	Left	**26**	**26**	2
	Right	10	10	11
Animal 2	Left	5	5	8
	Right	17	17	**27**

Fig. 1. (A) In both Experiments 1 and 2, participants completed two blocks of the reward revaluation task, followed by a recognition memory test. The task consisted of four phases: learning, re-learning, rest, and test. In the learning phase, participants made a series of two-stage decisions to earn treasure. They first chose an animal and then selected one of the animal's chests, which revealed the amount of treasure they would earn on that trial. During the relearning phase, participants only made second-stage choices between each animal's chests. In the revaluation condition, reward values during relearning were different from reward values in the learning phase, such that the optimal treasure chest choice for each animal changed. In the control condition, reward values did not change. After the relearning phase, participants in Experiment 1 completed a 1-min active rest phase where they were required to attend to the screen to perform a simple, non-cognitively demanding task. In Experiment 2, participants did not experience the rest phase, and instead proceeded directly from the relearning phase to the test phase. The test phase was designed to assess whether participants updated their first- and second-stage choice preferences based on the rewards they observed during relearning. In the test phase, participants made four first-stage choices without feedback, followed by eight second-stage choices without feedback. After completing two blocks (one in the revaluation condition, and one in the control condition) of the task, participants completed a test of recognition memory for the first-stage stimuli from the learning phase of the task. (B) Example reward values for a task block. During learning, participants experienced one set of treasure values (original values) and learned to navigate to the best chest (in bold) by first navigating to its corresponding animal. In the control condition, reward values remained unchanged during relearning, whereas in the revaluation condition, reward values changed so that the best chest (in bold) now belonged to a different animal.

the trial ended and participants lost five points.

After completing the learning phase, participants moved on to the re-learning phase of the task. In the re-learning phase, participants did not make first-stage choices. Instead, participants were told that they were traveling with a friend who would make first-stage choices for them. Participants were shown each of the two animals from the learning phase eight times (in a randomized order) and asked to select between their two chests. Importantly, the re-learning phase differed between the two blocks of the experiment. In one block of the task, participants experienced the *revaluation* condition, in which the amount of treasure in each chest changed between learning and re-learning so that the most valuable chest in the re-learning phase belonged to a *different* animal than in the original learning phase (Fig. 1B). The task also featured a *control* condition, where treasure amounts in the re-learning phase remained unchanged from the original learning phase. This was designed to control for the fact that participants may update their first-stage choices after relearning due to factors such as forgetting, rather than due to learning new reward values. Participants were not explicitly informed that rewards would or would not change at the start of this phase. The order of the revaluation and control blocks and the stimulus set (ocean or canyon) assigned to each condition were counterbalanced within each age group. As in the learning phase, participants had a time limit of 2 s to make each choice.

In Experiment 1 but not in Experiment 2, participants next completed an 'active rest' phase, during which we hypothesized they may 'replay' or reactivate the first-stage decisions, linking them via offline processing to the newly learned reward outcomes (Momennejad et al., 2018). During the rest phase, participants completed a 60-s task that was designed to ensure that they attended to the screen while being non-cognitively demanding. In the task, animated red dots moved slowly from the top to the bottom of the screen over approximately 4 s, and participants were instructed to click on them. Only one dot appeared on the screen at a time; new dots appeared 5 to 10 s after the previous dot was no longer present. All 119 participants in Experiment 1 included in the final sample missed fewer than four dots during the active rest phase.

After the rest phase, participants proceeded to 'test,' during which we assessed whether they updated their first-stage choices based on rewards experienced during relearning. In the test phase, participants made four first-stage choices, in which they saw the first-stage state (e.g., an image of an ocean or canyon, depending on what block they were in) and had to choose whether to go up or down. Participants were told to try to navigate to the animal with the most treasure. Unlike in the learning phase, here, to prevent continued learning, participants did not see any feedback, meaning they did not see which animal their choice led to. In addition, after making four first-stage choices, participants also made four second-stage choices starting from each of the two second-stage states (animals). Participants had 10 s to make each choice, but were not informed that they would have more time than in previous phases.

Finally, participants completed a surprise memory test for the first-stage state images they had seen during initial learning. We originally hypothesized that replanning would be facilitated by the reactivation of the first-stage states during the rest phase; we posited that such reactivation may also facilitate enhanced memory for the first-stage state images, such that participants who demonstrated the strongest replanning would also demonstrate the best memory for images from the learning phase, particularly within the revaluation condition. The images presented during learning were matched for memorability, such that images from the category used in each of the two blocks were

equivalently memorable (Lu et al., 2020). Scene images were each repeated twice during learning, such that 21 unique images were shown across the 42 learning trials in each block. To ensure attentiveness to the scene images, the learning task involved eight attention-check trials, during which a small, cartoon image of a robber was superimposed onto a first-stage scene image. Participants were told to press the spacebar if they spotted a robber in the environment; they were told that they failed to catch the robber if they did not respond. On robber trials, participants did not complete the two-stage decision task. The eight scene images seen on the attention-check trials were novel within-category scenes that were not seen on other trials, and not included in the recognition memory test.

During the recognition memory test, participants saw the 42 old images from the two learning blocks as well as 42 new images drawn from the same two scene categories. Participants were asked to determine if the presented image was 'Definitely New', 'Maybe New', 'Maybe Old', or 'Definitely Old.' Participants had 10 s to make each response and received no feedback. We report full recognition memory results in the supplement.

To ensure that child, adolescent, and adult participants fully comprehended the task, participants completed thorough, interactive instruction phases prior to its start. Task instructions featured child-friendly language and were presented both as text and via an audio recording. Participants could not advance each instruction screen until the corresponding audio track finished playing. Participants were also given the opportunity to practice the two-stage decision task, attention-check task, and the memory test before the real trials using a set of practice stimuli. At the end of each set of instructions, participants completed a set of True/False comprehension questions. There were four comprehension questions related to the learning task and two comprehension questions related to the memory test. After responding to a comprehension question, participants saw and listened to an explanation for the correct answer. Participants were required to answer all questions correctly to proceed in the experiment and were presented with the same question again if they made an error.

2.1.3. Analysis approach

We used the 'lme4' package (Bates et al., 2015) in R (R Core Team, 2021) to fit mixed-effects models to our data. Continuous variables were z-scored prior to model-fitting, and categorical variables were coded using sum contrasts. Age was treated continuously in all analyses. Trials in which participants failed to respond within the allotted time limit were excluded from analyses (Experiment 1: 0.01 % of trials for children aged 7–12, 0.01 % of trials for adolescents aged 13–17, 0.00 % of trials for adults; Experiment 2: 0.01 % of trials for children aged 7–12, 0.01 % of trials for adolescents aged 13–17, 0.00 % of trials for adults).

For all regression analyses, we began by fitting models that included random intercepts for each participant and random slopes for all fixed effects and their interactions for each participant (Barr et al., 2013). When models failed to converge, we pruned correlations between random slopes and intercepts, followed by interactions between random slopes, followed by random slopes themselves. Finally, when models with random intercepts only failed to converge (due to a lack of variation across subjects), we fit linear models using the 'stats' package in R with fixed effects only. We assessed the significance of fixed effects using Wald tests. We include the full specification for all models in the supplement.

3. Results

In Experiments 1 and 2, we asked whether participants across age leveraged a mental model of their environment to flexibly update their choice behavior when reward outcomes changed, with and without the opportunity for offline processing during rest, respectively. In our analyses, we first establish that participants learned to make optimal two-stage decisions during the learning phase. Next, we establish that when rewards in the environment changed, participants learned to make optimal second-stage choices in the relearning phase, and that they persist in making these new, optimal second-stage choices in the task's test phase. Finally, we assess whether participants effectively leverage structured knowledge to "replan" – meaning we ask whether they update their first-stage choices in the final test phase, even in the absence of direct experience.

3.1. Participants learned to make optimal two-stage decisions

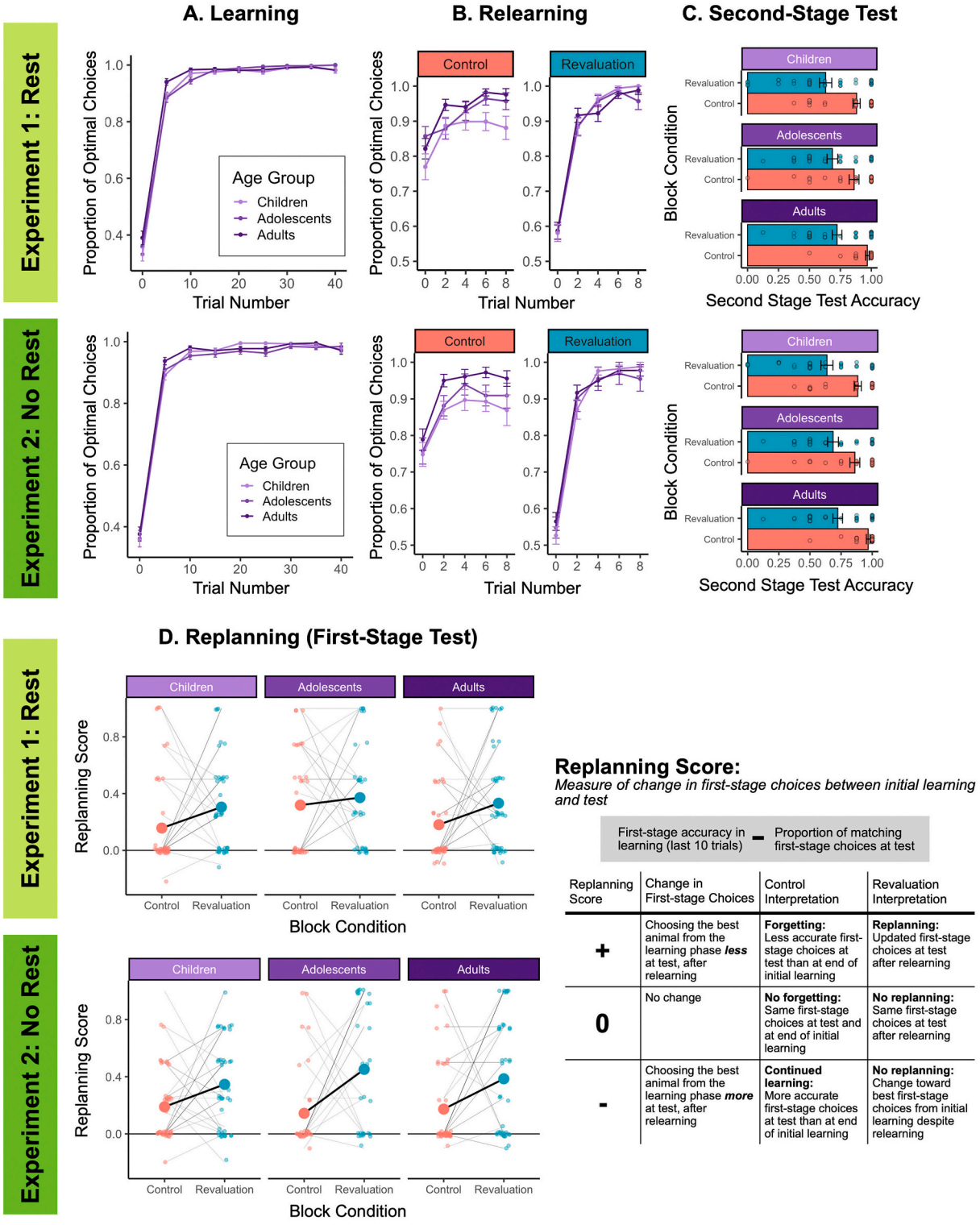
We first asked whether participants learned to make optimal two-stage choices over the course of the learning phase. We analyzed the influences of age, trial number, block condition, and their interactions on optimal choice via a mixed-effects logistic regression. Choices were considered 'optimal' (coded as 1) if participants made first- *and* second-stage decisions that would lead to the chest with the most treasure, and suboptimal otherwise (coded as 0). Across both experiments, we found that participants learned to make optimal choices across trials (Experiment 1 (E1): Log-Odds = 8.07, [7.06–9.08], z = 15.68, $p < .001$; Experiment 2 (E2): Log-Odds = 6.87 [6.03–7.72], z = 15.97, $p < .001$; Table S5) (Fig. 1B). We found no significant effect of age on optimal choices during learning (E1: Log-Odds = −0.33 [−1.48–0.82], z = −0.57, $p = .57$; E2: Log-Odds = −0.12 [−1.09–0.85], z = −0.24, $p = .81$; Table S5), suggesting that participants across our entire age range successfully learned to make multi-step decisions to navigate to the most rewarding treasure chest. Across age, participants learned to select the chest with the most treasure very reliably, achieving high accuracy in the last ten trials of learning (E1: children 98.9 % (SE = 0.4 %), adolescents 99.9 % (SE = 0.1 %), adults 98.9 % (SE = 0.4 %); E2: children 98.6 % (SE = 0.6 %), adolescents 98.5 % (SE = 0.6 %), adults 99.0 % (SE = 0.4 %)). No other effects or interactions were significant (see Supplement for full results).

3.2. Successful re-learning of new reward values

After confirming that participants learned to make optimal two-stage choices, we next assessed whether they updated their beliefs about the most rewarding treasure chests when they experienced new treasure amounts in the revaluation task condition. To do so, we examined the influence of block condition, age, and number of exposures to each second-stage state (adjusted trial number) on optimal choice during relearning. In line with the task manipulation, we found a significant effect of block condition (E1: Log-Odds = −0.40 [−0.67 to −0.13], $z = -2.92$, $p = .003$; E2: Log-Odds = −0.71 [−1.08 to −0.35], $z = -3.85$, $p < .001$; Table S6), such that participants made fewer optimal choices in the revaluation condition in which treasure amounts differed from those in the original learning phase versus in the control condition, in which treasure amounts remained the same. However, across trials, participants learned the new reward values, making more optimal choices across exposures (E1: Log-Odds = 1.53 [1.30–1.75], $z = 13.32$, $p < .001$; E2: Log-Odds = 1.95 [1.66–2.25], $z = 12.93$, $p < .001$; Table S6), particularly in the revaluation condition compared to the control condition (E1: Log-Odds = −0.76 [−0.99 to −0.54], $z = -6.60$, $p < .001$; E2: Log-Odds = −0.95 [−1.24 to −0.67], $z = -6.61$, $p < .001$; Table S6). There was no significant effect of age on re-learning performance (E1: Log-Odds = 0.10 [−0.10–0.31], $z = 1.01$, $p = .31$; E2: Log-Odds = 0.23 [−0.04–0.51], $z = 1.69$, $p = .09$; Table S6). Surprisingly however, younger participants were less optimal in the control condition compared to older participants (marginally in E2) (E1: Log-Odds = 0.32 [0.12–0.51], $z = 3.15$, $p = .001$; E2: Log-Odds = 0.26 [0.01–0.53], $z = 1.89$, $p = .059$; Table S6), and demonstrated less improvement with experience in Experiment 1 (E1: Log-Odds = 0.19 [0.02–0.36], $z = 2.21$, $p = .03$; E2: Log-Odds = 0.17 [−0.04–0.38], $z = 1.59$, $p = .11$; Table S6) (Fig. 2B), potentially reflecting boredom or disengagement. This effect persisted regardless of whether participants encountered the control condition before or after the revaluation condition (see Table S7 in the

Supplement). Nevertheless, across age, accuracy at the end of re-learning was still high (Fig. 2B), indicating that across blocks, participants successfully learned the treasure amounts in the animals' chests.

To further confirm that participants learned the second-stage reward values during relearning, we next examined whether participants made optimal second-stage choices without feedback in the test phase (Fig. 2C). At test, when presented with each animal, participants chose the more rewarding treasure chest at above-chance levels in both the revaluation and control conditions (E1: revaluation 68.0 % (SE = 2.5 %), control 90.5 % (SE = 1.7 %); E2: revaluation 71.7 % (SE = 2.5 %), control 89.1 % (SE = 1.7 %)), confirming that they successfully retained and used 'relearned' chest values to guide their choices. As in relearning, there was a significant effect of block condition on second-stage test accuracy such that participants were more accurate in the control



Replanning Score:
Measure of change in first-stage choices between initial learning and test

First-stage accuracy in learning (last 10 trials) − Proportion of matching first-stage choices at test

Replanning Score	Change in First-stage Choices	Control Interpretation	Revaluation Interpretation
+	Choosing the best animal from the learning phase ***less*** at test, after relearning	**Forgetting:** Less accurate first-stage choices at test than at end of initial learning	**Replanning:** Updated first-stage choices at test after relearning
0	No change	**No forgetting:** Same first-stage choices at test and at end of initial learning	**No replanning:** Same first-stage choices at test after relearning
-	Choosing the best animal from the learning phase ***more*** at test, after relearning	**Continued learning:** More accurate first-stage choices at test than at end of initial learning	**No replanning:** Change toward best first-stage choices from initial learning despite relearning

(caption on next page)

Fig. 2. Results from Experiment 1 and Experiment 2. All analyses treated age continuously; age is binned into groups for visualization purposes. (A) Participants learned to navigate to the most rewarding treasure chest (by making optimal first- and second-stage choices) over the course of the learning phase in both experiments. Performance improved across trials and did not significantly vary with age. For visualization, the proportion of optimal choices is calculated over bins of 5 trials, and error bars show standard errors of participant means. (B) In the relearning phase, participants made second-stage choices only and observed the associated rewards. In the revaluation condition, in which reward values changed during relearning, participants made fewer optimal choices on early trials, but rapidly learned to respond optimally, as they gained additional experience with the new reward values. Younger participants performed slightly worse than older participants. For visualization, the proportion of optimal choices is calculated over bins of 2 trials, and error bars reflect standard errors across participant means. (C) Participants updated their second-stage choices at test based on reward experience from relearning and performed above chance in both conditions. As in relearning, older participants made more accurate choices, and participants across age made more accurate choices in the control condition compared to the revaluation condition. Points indicate individual accuracies and bar widths indicate age-group means. Error bars reflect standard errors across participant means. (D) In both experiments, participants across age updated their first-stage choices (replanned) more in the revaluation condition compared to the control condition. Replanning scores index changes in first-stage choices between the last 10 trials of the initial learning phase and all four test phase trials. Higher replanning scores in the revaluation condition relative to the control condition indicate that participants changed their first-stage choices more when they experienced new second-stage reward values during relearning. Replanning scores did not vary across age or across experiments. Smaller points indicate individuals' replanning indices, while the larger points indicate age-group means.

condition relative to the revaluation condition (E1: Estimate = 0.11 [0.08–0.14], $z = 7.93$, $p < .001$; E2: Estimate = 0.09 [0.06–0.11], $z = 6.27$, $p < .001$; Table S7). This difference may reflect forgetting of the new reward values from the re-learning phase (but remembering the initially learned values from the original, longer learning phase) or a belief that the reward environment at test had reverted to the original learning environment. However, participants' above-chance performance on second-stage test trials suggests that they generally understood that their relearning experience should inform their choices at test.

Additionally, across conditions, older participants' second-stage choices were slightly more accurate than younger participants (E1: Estimate = 0.05 [0.02–0.08], $z = 2.95$, $p = .004$; Children: 75.6 % (SE = 3.2 %), Adolescents: 77.3 % (SE = 3.1 %), Adults: 84.5 % (SE = 2.4 %); E2: Estimate = 0.06 [0.03–0.09], $z = 3.83$, $p < .001$; Children: 75.3 % (SE = 2.5 %), Adolescents: 75.0 % (SE = 3.6 %), Adults: 89.0 % (SE = 2.2 %); Table S7). There was no significant interaction between age and block condition on second-stage accuracy at test (E1: Estimate = 0.00 [−0.03–0.02], z = 0.35, $p = .73$; E2: Estimate = 0.00 [−0.02–0.03], z = 0.22, $p = .83$; Table S7), further supporting the idea that age differences in the control condition during relearning were due to the control condition being less engaging rather than differences in learning the reward values.

3.3. Evidence for revaluation of first-stage choices

Finally, we investigated our main question of interest — whether participants used their knowledge of the structure of the environment to update their multi-step plans in the absence of direct experience. To do so, we examined participants' first-stage choices during the test phase, in which they received no feedback. In the revaluation condition, the best treasure chest in the relearning phase was associated with a *different* animal than it was during the original learning phase. We hypothesized that if participants integrated new reward values with their mental model of the task structure, then they would change their first-stage choices in the revaluation condition but not the control condition. Further, we hypothesized that if revaluation depended on reactivating first-stage choices during rest, participants would change their first-stage choices in the revaluation condition to a greater degree in Experiment 1, which included rest, versus Experiment 2, which did not.

We computed a 'replanning score' for each participant for each task block, which indexes the extent to which they made different first-stage choices in the test phase versus the initial learning phase. Replanning was computed by taking participants' mean first-stage choice accuracy on the last 10 trials of the task's original learning phase and subtracting the proportion of first-stage test trials on which they made the original, best first-stage choice (see Fig. 2D). This means that if participants showed perfect replanning in the revaluation condition (i.e., if they performed perfectly at the end of the learning phase and then reliably switched to the *other* first-stage choice at the test phase), they would have a replanning score of 1, indicating that they always selected the original, best first-stage choice during learning and never selected the original, best first-stage choice at test. If they showed no replanning, they would have a replanning score of 0, indicating that they made the *same* first-stage choices during the initial learning and test phases. Rather than replanning, a change in performance from learning to test might also occur due to forgetting. The control condition controls for this possibility; In the control condition, a positive replanning score measures this effect because in the control condition, adaptive planning does not favor switching choices. A larger replanning score (more switching) in the revaluation relative to the control condition, in turn, indicates successful replanning. For our main analyses, we rely on the "replanning score" measure, but we additionally report raw accuracy for first-stage test trials (Fig. S1) as well as an analysis of the effect of test trial number on first-stage choices (Table S1 and Fig. S2) in the Supplement.

In line with our hypothesis, in Experiment 1, we found a significant effect of revaluation condition on replanning score (Estimate = −0.06 [−0.10 to −0.02], $z = -2.68$, $p = .008$, Table S8), indicating that participants updated their first-stage choices based on their experiences with the second-stage rewards during relearning. We found no significant effect of age (Estimate = 0.02 [−0.02–0.07], $z = 0.94$, $p = .35$, Table S8) or interaction between age and block condition on replanning score (Estimate = 0.00 [−0.04–0.04], $z = 0.01$, $p = .99$, see Fig. 2D and Table S8), suggesting that participants across our age range similarly replanned more in the revaluation condition compared to the control condition. In Experiment 2, we similarly found a significant effect of revaluation condition on replanning score (Estimate = −0.11 [−0.15 to −0.06], $z = -4.82$, $p < .001$, Table S8), indicating that participants 'replanned' even without a rest period. As in Experiment 1, the effect of block condition on replanning did not vary with age (Estimate = −0.01 [−0.04–0.05], $z = 0.22$, $p = .82$, Table S8). However, we observed a high degree of variability in replanning across individuals (see Fig. 2D), such that some participants did not replan in the revaluation condition, or even replanned more in the control condition compared to the revaluation condition. To provide further insight into this variability, we include an analysis of participants' replanning as a function of both age and their performance on second-stage test trials in the Supplement (see Fig. S3). Briefly, in this supplemental analysis, we found that participants who demonstrated greater accuracy on second-stage test trials also demonstrated more flexible replanning, and that this effect remained consistent across age.

Finally, we directly examined whether the opportunity for rest facilitated non-local learning by analyzing data from Experiments 1 and 2 together. Critically, we did not observe evidence that rest influenced replanning; the block condition x experiment interaction on replanning score was not significant (Estimate = −0.02 [−0.06–0.01], $z = -1.52$, $p = .13$, Table S9), with the estimated trend in the direction against the hypothesis that rest facilitates planning. In addition, though we initially hypothesized that children may benefit more from the opportunity for

offline replay than adults, we did not find a significant block condition x experiment x age interaction, (Estimate = −0.01 [0.04–0.02], $z = 0.72$, $p = .47$, Table S9), indicating that we did not observe evidence that rest facilitated replanning to a greater extent in younger participants.

4. Interim discussion

In Experiments 1 and 2, we assessed whether participants spanning childhood to early adulthood could integrate newly learned reward values with structured knowledge to update their behavior in a reward revaluation task. We further asked whether this ability depended upon the opportunity for offline integration during a rest phase. We found that children and adolescents were able to leverage knowledge of the task's transition structure to flexibly update their choice behavior when reward values changed. While we initially hypothesized that behavioral flexibility would be supported by offline replay, we found that participants demonstrated flexible choice behavior regardless of whether they had the opportunity for offline processing during rest. Rest did not significantly influence the extent to which children, adolescents, or adults replanned.

One possible explanation for these findings is that the 60-s rest period included in Experiment 1 was not necessary for participants to engage in offline processing or replay. Instead, participants may already have reactivated the relevant first-stage choice state immediately after experiencing second-stage rewards in the relearning phase (Foster & Wilson, 2006; Liu, Mattar, et al., 2021; Singer et al., 2013; Singer & Frank, 2009; Wimmer et al., 2023; Wimmer & Shohamy, 2012). Indeed, in past studies where the inclusion of a rest phase seemed to influence replanning, the relearning phase was performed under cognitive load (Gershman et al., 2014). This suggests that cognitive load may have reduced the opportunity for offline processing within the relearning phase itself and heightened the importance of the rest phase. Additionally, past work that demonstrated a link between replanning and neural reactivation during the rest phase did not manipulate rest (Momennejad et al., 2018). Therefore, it is possible that the reactivation observed during that period did not *causally* increase replanning — instead, participants who demonstrated the greatest reactivation of first-stage states during rest may have also reactivated those states to a greater extent during the relearning phase itself, facilitating re-planning. In the present study, engaging in on-task and offline replay could have both enabled effective replanning. However, in cases where reward values change throughout the task, and not solely before periods of rest, on-task replay may facilitate increased behavioral flexibility over offline replay (Eldar et al., 2020; Ólafsdóttir et al., 2017). Future work could further examine developmental differences in both on-task and offline replay, for example by adding cognitive load manipulation (Gershman et al., 2014) to the current paradigm to reduce replay during the relearning phase, or by using neural decoding methods (Kurth-Nelson et al., 2016; Liu, Dolan, et al., 2021; Schuck & Niv, 2019) to detect online and offline replay in a developmental sample.

Another possible explanation for our results is that participants *do* rely on prospective, model-based planning in the task's test phase. While prior work has found that model-based planning increases from childhood to adulthood, those studies have largely used tasks with greater cognitive demands (Decker et al., 2016; Ma et al., 2022; Nussenbaum, Scheuplein, et al., 2020; Smid, Kool, et al., 2023). For example, the present reward revaluation task includes deterministic transitions between states and deterministic reward outcomes, whereas other tasks assessing model-based planning (e.g., 'the two-step task' (Daw et al., 2011)) involve probabilistic transitions between states and probabilistic rewards. Therefore, it may be less computationally demanding for participants to simulate multi-step decisions in the current task. This would account for the lack of age-related differences observed here, both during the initial learning phase, and in replanning behavior. It is important to note, however, that more children were excluded from our analysis due to poor initial learning compared to adolescents and adults. This means that the children included in our sample were all capable of learning to make two-stage decisions to lead to reward. This inclusion criterion, while necessary for evaluating replanning behavior, may mean that the younger participants in our sample were generally better reward learners and more capable of replanning relative to the general population of their same-aged peers.

A third possibility is that children and adolescents leverage predictive representations — like the successor representation — to behave flexibly without the need for either offline replay or iterative model-based planning. In the initial learning phase of our task, participants may have learned the probabilities of ending up in each second-stage state following each first-stage choice, or in other words, they may have cached a representation of each first-stage state's expected successors. After re-learning new reward values, participants may have leveraged these previously learned probabilities to rapidly assess the value of each first-stage choice (Dayan, 1993; Gershman, 2018; Momennejad et al., 2017; Russek et al., 2017). For example, a participant may have learned that if they chose the octopus, they then typically chose the left treasure chest, thereby forming a more temporally abstract or 'predictive' representation of the octopus that takes into account the likely transition to the left chest. When the reward within the left chest changed during the revaluation phase, they may have then automatically updated their value representation of the octopus. In our revaluation task, using a predictive representation like the SR could support successful replanning because, although the final 'best' chest in the revaluation condition was always associated with the initial, worse, first-stage choice, it was also always located in that animal's 'better' treasure chest that would have been chosen after that first-stage choice during learning (Fig. 1). Thus, participants could rely on cached knowledge of likely state transitions within the task to update first-stage value representations.

While this is an intriguing possibility, to the best of our knowledge, no prior studies have examined whether children harness predictive representations like the SR to guide their decisions. Successful revaluation in Experiments 1 and 2 could be attributed to the use of either model-based or SR-based strategies; our data cannot distinguish between the two. Thus, in Experiment 3, we sought to directly investigate whether children and adolescents use the SR for flexible decision making, and whether use of the SR changes across development. We assessed use of the SR by adapting a multi-trial reinforcement learning task from recent adult work that was designed to distinguish between model-free, model-based, and SR-based decision strategies (Kahn & Daw, 2025). We hypothesized that use of SR-based strategies might be evident from childhood; using the SR to guide decisions does not require the computationally expensive simulation of multi-step outcomes at choice time (Gershman, 2018), and thus may be a more effective decision strategy for children and adolescents, who are good at learning the statistical regularities of their environments (Forest, Schlichting, et al., 2023).

In Experiment 3, we were also interested in whether children and adolescents rationally trade off between the use of different decision strategies. While the SR can enable efficient and flexible choice, it is only useful when the transition structure of the environment is relatively stable. Indeed, this is one reason why participants may not have been able to rely on it in other, more dynamic tasks used to assess model-based planning over development (Daw et al., 2011; Decker et al., 2016; Piray & Daw, 2021). If the SR that has been learned captures environmental statistics that are no longer relevant, use of the SR would be maladaptive. In such cases, only the use of model-based planning would allow for full behavioral flexibility. In prior work, adults adaptively reduced reliance on the SR when its underlying assumptions were violated (Kahn & Daw, 2025). Here, we hypothesize that children may be less able to rationally trade off between SR-based and model-based strategies as compared to adults. This idea is supported by evidence from prior work suggesting that children may demonstrate reduced "meta-control," such that they do not arbitrate between different

decision strategies as effectively as adults (Bolenz & Eppinger, 2022; Smid, Ganesan, et al., 2023; Smid, Kool, et al., 2023).

5. Experiment 3

5.1. Methods

5.1.1. Participants

152 participants between the ages of 8–22 years completed Experiment 3 online, remotely and asynchronously, and were included in our analyses. Of these 152 participants, 50 were children (8.02–12.97 years; Mean age = 10.46, n = 25 females), 51 were adolescents (13.03–17.92 years; Mean age = 15.30, n = 26 females), and 51 were adults (18.15–22.69 years; Mean age = 20.59, n = 26 females). An additional 9 participants completed the study but were excluded from all analyses for making 3 or more errors on the task comprehension questions (n = 6 Children, n = 2 Adolescents, n = 1 Adult). All participants interacted with their browser window fewer than 20 times during the study session, and all participants responded on more than 95 % of the tasks' 200 trials. Sample size and exclusion criteria were defined a priori as in Experiments 1 and 2.

Participants were recruited for this experiment in the same manner as for Experiments 1 and 2. According to self- or parental-report, 43.4 % of participants were White, 30.3 % were Asian, 13.2 % were Black, and 13.2 % were two or more races. 15.8 % of participants were Hispanic.

5.1.2. Experimental procedure

We adapted a planning task used in a recent adult study (Kahn & Daw, 2025) that was designed to distinguish between use of model-based (MB) planning and use of the successor representation (SR). This task provides a dynamic, trial-by-trial measure of the use of these learning strategies, allowing us to robustly assess individual differences in their usage and to look at flexible, within-individual arbitration between them.

As in Experiments 1 and 2, the online task was programmed in jsPsych and hosted on Pavlovia. Participants completed a thorough, interactive instruction phase prior to the start of the task with child-friendly language and audio recordings. Participants also completed a set of True/False comprehension questions at the end of each instruction block, and repeated both the instructions and comprehension questions up to three times if they answered them incorrectly. Participants also practiced the task using a set of practice stimuli before beginning the real task.

In the task, participants sailed to different islands to collect treasure from island shops. Participants were instructed to collect as much treasure as possible, and were paid a bonus based on the amount that they collected. Unlike in Experiments 1 and 2, here, treasure amounts were binary (i.e., 1 or 0) on each trial. In the task, participants could visit two different islands, each of which had two differently colored shops. Upon reaching a shop, participants would receive treasure with a shop-specific reward probability. Reward probabilities for all shops changed across blocks, which were not signaled to participants and comprised between 16 and 24 trials (Fig. 3B). Participants were told that if the shopkeepers had been successful recently, they would share their treasure with them. They were also told that the fortune of the shopkeepers may change and that a shop that provided treasure often early on in the task may later provide treasure only rarely.

Importantly, there were two types of trials during the task: traversal trials and non-traversal trials (Fig. 3A). These trials were presented alternatingly to the participant, and each participant completed 200 of each type of trial. In traversal trials, participants made two-stage decisions, as they did in the learning phase of Experiments 1 and 2. In the first stage, they selected between two islands by pressing the left or right arrow keys on a standard keyboard. After sailing to the selected island, they then made a second-stage decision about which shop to visit, again using the left or right arrow keys. After selecting a shop, they saw the outcome of their choice, meaning that they saw that they either received or did not receive treasure. Participants pressed the spacebar to collect treasure. For traversal trials, participants were allowed up to 10 s to make each choice. If they did not make a choice within the time limit, they were issued a warning and did not earn any treasure.

In non-traversal trials, participants did not navigate to an island or shop. Instead, they were transported to a randomly selected shop (without the island visible) and were told that they had arrived at the shop on a cloudy day. They then saw, and pressed the spacebar to receive, the reward (treasure or no treasure) for that shop. There was no response time limit on non-traversal trials. This mirrors the relearning phase of Experiments 1 and 2, where participants were given the opportunity to learn about reward values without making a first-stage choice. In this experiment, however, traversal and non-traversal trials alternated, allowing for many repeated measures of 'non-local' learning — here, the influence of non-traversal rewards on subsequent island choices.

This task was additionally designed to test whether individual participants flexibly trade off between MB and SR strategies (Kahn & Daw, 2025). To test for adaptive arbitration between strategies, shop reward probabilities changed between blocks in two different ways (Fig. 3B). After *congruent* block changes, the most rewarding shop was now located on a different island as on the previous block, but the best shop on each island remained the same. After congruent block changes, the previously learned SR was still useful because the most likely 'successor' shop for each island should remain consistent (e.g., a participant who frequently chose Shop A on Island 1 should continue to choose Shop A on that island). After *incongruent* block changes, the most rewarding shop was now located on a different island *and* the most rewarding shop on each island also changed. After incongruent block changes, use of the previously learned SR was maladaptive, because it reflects transition probabilities based on a now outdated choice policy (e.g., a participant who frequently chose Shop A on Island 1 should now change their policy to choose Shop B on that island). Participants experienced 21 unsignaled block changes over the course of the task, of which 15 were congruent and 6 were incongruent. The order of congruent and incongruent block changes was randomized across participants. By examining participants' reliance on MB and SR strategies across block types, we could test the extent to which they flexibly arbitrated between learning strategies based on the predictability of the environment's transition structure.

5.1.3. Analysis approach

We fit mixed-effects models for Experiment 3 following the same approach as for Experiments 1 and 2.

5.1.4. Mixture of agents reinforcement-learning model

In Experiment 3, we additionally characterized participant choice behavior with the 'Mixture of Agents' reinforcement-learning model used in the original adult study (Kahn & Daw, 2025). Briefly, each 'agent' learns the value of each island via a different learning algorithm (described in detail below). The model then combines these values to determine which island to select on each trial; the weights the model assigns to each learning algorithm are determined by three separate inverse temperature parameters $(\beta_{MF}, \beta_{MB}, \beta_{SR})$ that are fitted to each individuals' choices. The values of inverse temperature parameters therefore reflect the contributions of MF-, MB-, and SR-based learning to each participant's choices.

In the learning model, all agents similarly update their estimate of the value of each shop $V(shop_i)$ after observing the trial's reward outcome R_t:

$$V(shop) \leftarrow (1-\alpha)V(shop) + \alpha R_t$$

The model includes separate learning rates (α) for traversal and non-traversal trials, to capture potential differences in learning about the shops after choosing to visit them (on traversal trials) versus passively arriving at them (on non-traversal trials). On traversal trials, choices

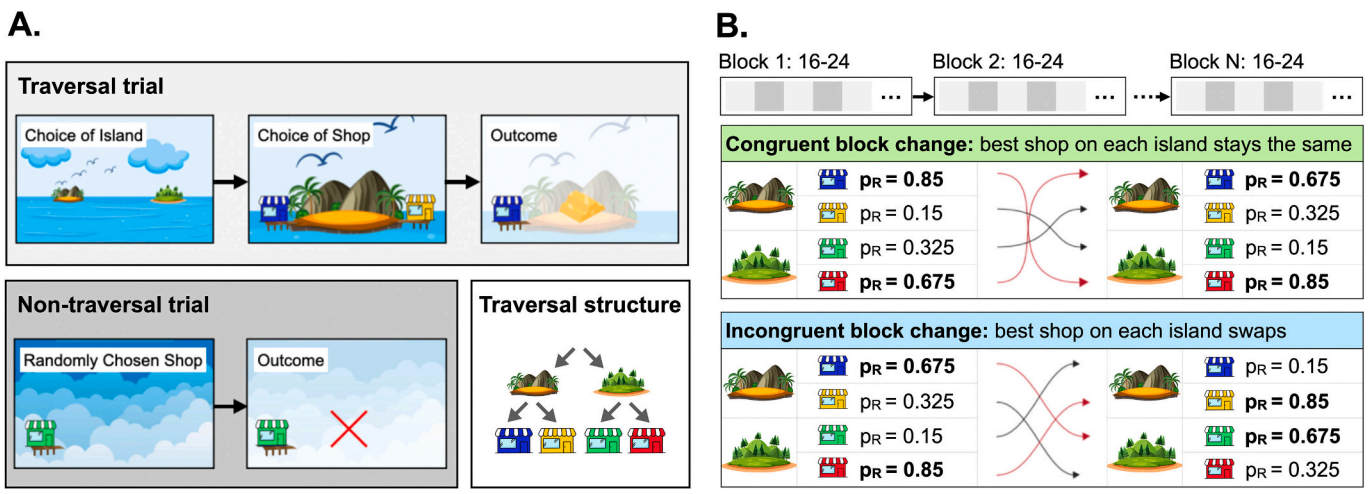


Fig. 3. (A) In the task, participants sailed to islands in order to collect treasure from shops on those islands. There were two islands to choose between and two shops on each island for a total of four shops. Each shop had a probability of providing a binary reward that changed over the course of the task. The task consisted of traversal trials and non-traversal trials. In traversal trials, participants first chose an island to sail to, then chose a shop on that island, and lastly saw the reward outcome associated with that shop. In non-traversal trials, participants did not choose an island. Instead, they were told that they ended up at a random shop on a cloudy day, and then they saw the reward outcome associated with that shop. (B) Each block of the task consisted of 16–24 trials that alternated between traversal and non-traversal trials. Reward probabilities of all shops changed at the start of each new block. These block changes were not explicitly signaled to participants. Importantly, two types of block changes occurred throughout the task. In congruent block changes, the best island to choose switched, but the best shop on each island remained the same. In incongruent block changes, both the best island and the best shop on each island switched. After congruent block changes, but not after incongruent block changes, evaluating the new rewards based on the policy from the previous block results in optimal island choices. The best shop on each island is indicated in the figure above via bolded text.

between the shops were modeled with a softmax decision rule, such that:

$$P(shop_t = shop) \propto exp\left(\beta_{shop} V(shop) + \beta_{sticky_s} LastChosen(shop)\right)$$

where β_{shop} is an additional inverse temperature parameter that captures the extent to which participants' shop choices were value-driven, β_{sticky_s} is a stickiness parameter that captures perseverative tendencies, and $LastChosen(shop) = 1$ if the shop was the most recently selected shop on the current island, and 0 otherwise.

The MF agent updates its estimate of the value of the selected island on traversal trials, taking into account their previous belief about the value of the island, as well as the value of the shop that it selected there, such that:

$$V_{MF}(island) \leftarrow (1-\alpha) V_{MF}(island) + \alpha V_{MF}(shop)$$

The MB agent computes the value of each island simply by taking the maximum of the values of its two shops.

$$V_{MB}(island) = max(V(shop_1), V(shop_2))$$

The SR agent learns a matrix, M, of future state occupancies, that captures the likelihood of transitioning to each shop from each island. On traversal trials, M is updated, such that:

$$M[island, shop] = \alpha_M 1_{shop=c} + (1-\alpha_M) M[island, shop]$$

where α_M is a free parameter that reflects the learning rate of the M matrix and c reflects the choice the participant made on that trial. The SR agent then computes the island values by taking the product of M and R, where R is a vector of estimated shop values:

$$V_{SR}(island) = MR$$

Finally, as noted above, the choices between islands are modeled as a probabilistic decision between them, with the values of the islands determined by weighting the island values computed by the three learning agents:

$$P(island_t = i) \propto exp\Big(\beta_{MF} V_{MF}(i) + \beta_{MB} V_{MB}(i) + \beta_{SR} V_{SR}(i) + \beta_{sticky_I} LastChosen(island)\Big)$$

where β_{sticky_I} is a stickiness parameter that captures perseverative tendencies in each participant's island choices.

Free parameters of the model were estimated using an expectation-maximization algorithm (Huys et al., 2011) implemented in Julia. Subject-level parameters were modeled as arising from population-level Gaussian distributions over subjects, where each Gaussian distribution was parameterized by its mean and variance. To estimate developmental differences in parameters, we also included an age covariate, which allowed the population-level mean to vary linearly with age. We include parameter recoverability analyses in the Supplement, which demonstrate the model's ability to accurately estimate age-related change in the inverse temperature parameters of interest.

6. Results

6.1. Participants leveraged mental models to guide choice

To use reward experienced on non-traversal trials to guide their subsequent island choices, participants must leverage a mental model that links the observed shop to the island on which it is located — a pure 'model-free' (MF) learner would not learn anything about the value of the islands from non-traversal trials. Both SR-based and MB decision strategies, however, enable this kind of flexible learning from non-traversal trial reward outcomes, using different forms of mental models. Agents using a fully MB strategy would compute the value of each island by conducting step-by-step forward simulations of the two sequential choices they could make and the reward outcome they would likely experience. Thus, an MB agent would assess the value of each island as equivalent to the value of the most rewarding shop on that island. The SR offers a simplified model linking islands to outcomes: Rather than relying on step-by-step simulation, an SR agent would assess

the value of an island based on the shop it expected to visit after arriving there, learned from its prior experiences. Learners using either strategy should be more likely to choose to visit an island after experiencing a reward from a shop on that island during a non-traversal trial.

We first assessed whether participants across age leveraged structured knowledge to support their decision making by examining how the reward they experienced on non-traversal trials affected their subsequent traversal-trial island choice. Importantly, reward probabilities in the task were matched across islands, such that in every block, shops sampled from both islands were overall equally likely to yield reward. We found that after experiencing reward from a particular shop on a non-traversal trial, participants were more likely to choose to visit the island where that shop was located on the next traversal trial, relative to trials where they did not receive reward (Log-Odds = 0.19 [0.15–0.22], $z = 10.07$, $p < .001$, Table S10). This effect did not significantly interact with age (Log-Odds = 0.01 [−0.02–0.05], $z = 0.74$, $p = .46$, Table S10), suggesting that participants across our age range leveraged structured knowledge to guide choice.

6.2. Unique signatures of MB and SR-based decision strategies

As previously described, both MB and SR-based strategies could, in theory, enable participants to use rewards experienced on non-traversal trials to guide their subsequent island choices. However, MB and SR-based decision strategies have unique behavioral signatures in this task. The extent to which a reward from Shop A influences an MB-learner's island choice will depend on whether the *other* shop on that island (in this case, Shop B) was recently rewarding. If the MB-learner recently experienced a reward from Shop B, then they would already be likely to choose Island 1, regardless of the reward they experienced from Shop A. The extent to which a reward from Shop A influences an SR-learner's island choice will depend not on their beliefs about the other shops, but rather on their beliefs about the likely transitions they will experience in the environment. These transitions depend on the learner's own policy, and specifically, how often the learner selects Shop A on Island 1. At an extreme, if an SR-learner always picks Shop B on Island 1, then non-traversal rewards from Shop A should not influence their choices, because Shop A is an unlikely successor to Island 1, and therefore not part of their Island 1 predictive representation. In other words, for an SR agent, the influence of the non-traversal reward outcome on their subsequent island choice should be modulated by their previous shop choices on that island. These distinct behavioral signatures enable measurement of peoples' use of MB and SR strategies within this task, on a trial-by-trial basis.

We looked for behavioral signatures of each decision strategy by examining how the most recently experienced reward from the neighboring shop and the most recent shop choice made on the island modulate the influence of the non-traversal reward outcomes on subsequent island choices. As in our prior analysis, we continued to observe a significant effect of non-traversal trial reward outcomes on subsequent island choices (Log-Odds = 0.25 [0.21–0.29], $z = 11.53$, $p < .001$, Table S11). Here, we also observed a reward x neighboring shop reward interaction effect (Log-Odds = −0.11 [−0.14 to −0.08], $z = -6.78$, $p < .001$, Table S11), such that participants showed a greater influence of non-traversal trial rewards on subsequent island choices when the neighboring shop on the island was previously not rewarding (Fig. 4A), in line with the behavior of a MB agent. We further observed a reward x prior shop choice interaction effect (Log-Odds = 0.12 [0.08–0.15], $z = 6.72$, $p < .001$, Table S11), such that participants showed a greater influence of non-traversal trial rewards on subsequent island choices when they had previously chosen to visit that shop on the island (Fig. 4B), in line with the behavior of an SR-based agent. Thus, participants demonstrated evidence of using both step-by-step simulation and cached, simplified predictive representations to leverage structured knowledge for value-guided choice.

We initially hypothesized that MB learning would increase with age, whereas children would show early-emerging reliance on an SR-based decision strategy. Here, however, we did not observe evidence for significant age-related changes in these behavioral signatures of the two decision strategies (SR effect: Log-Odds = 0.03 [−0.01–0.06], $z = 1.67$, $p = .09$; MB effect: Log-Odds = −0.03 [−0.06–0.01], $z = -1.61$, $p = .11$, Table S11). In addition, when we restricted our analyses to children only (8.02–12.97 years, $n = 50$), we continued to observe a significant effect of non-traversal reward outcomes on subsequent island choices (Log-Odds = 0.24 [0.18–0.31], $z = 7.40$, $p < .001$, Table S12), as well as significant evidence of both MB behavior (reward x neighboring shop reward interaction: Log-Odds = −0.10 [−0.15 to −0.05], $z = -3.69$, $p < .001$, Table S12) and SR-based behavior (reward x prior shop choice interaction: Log-Odds = 0.09 [0.03–0.15], $z = 3.14$, $p = .002$, Table S12). These analyses suggest that in a similar manner as adults, children harnessed multiple learning strategies that leverage structured knowledge to make decisions.

6.3. Divergent developmental trajectories of different decision strategies

While the simpler regression analyses did not reveal evidence for age-related change in choice strategies, they rely on coarse approximations of strategy that only take into account participants' choices and

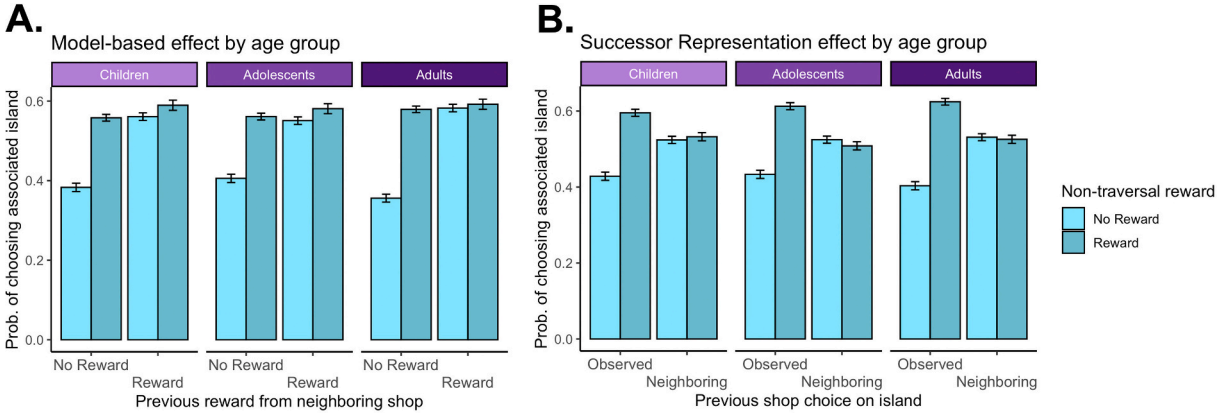


Fig. 4. Participants across age exhibited signatures of model-based and SR-based behavior: Participants demonstrated a stronger effect of non-traversal trial rewards on subsequent island choices both when they had previously not received a reward from the neighboring island shop (model-based signature; see main text) and when they had previously chosen the island shop on the island (SR-based signature; see main text). There were no significant age-related differences in either of these behavioral signatures. Age is binned into groups for visualization purposes. Bar heights indicate age-group means, while error bars reflect standard errors of participant means.

experienced rewards from the most recent trials. To account for participants' full history of learning experiences across trials — and individual differences in how they learned from rewards — we further characterized participant choices with the 'Mixture of Agents' reinforcement learning model used in the original adult study (Kahn & Daw, 2025). Prior developmental work examining the use of structured knowledge to guide decision making has taken a similar approach to characterize the extent to which behavior reflects the contributions of MF- and MB-learning agents (Decker et al., 2016; Nussenbaum, Scheuplein, et al., 2020; Potter et al., 2017). The model used here additionally characterizes the contributions of an SR-based learning agent. Briefly, each 'agent' learns the value of each island via a different learning algorithm (described in detail in the methods). The model then combines these values to determine which island to select on each trial; the weights the model assigns to each learning algorithm are determined by three separate inverse temperature parameters ($\beta_{MF}, \beta_{MB}, \beta_{SR}$) that are fitted to each individuals' choices. The values of inverse temperature parameters therefore reflect the contributions of MF, MB, and SR-based learning to each participant's choices. To estimate developmental differences in model parameters, we fit this model hierarchically, and allowed the population-level mean of all parameters to vary linearly with age.

In line with our original regression analyses, we found significant evidence for contributions from both MB and SR-based learning mechanisms, as reflected in population-level β_{MB} and β_{SR} estimates that were significantly greater than 0 (β_{MB}: mean = 0.59, $SE = 0.10$, $p < .001$; β_{SR}: mean = 0.41, $SE = 0.07$, $p < .001$, see Fig. 5A). Here, we also found a significant influence of MF learning (β_{MF}: mean = 0.40, $SE = 0.06$, $p < .001$). Together, these estimates indicate that participants used a combination of multiple learning strategies to guide their decisions, in line with prior work (Daw et al., 2011; Decker et al., 2016; Kahn & Daw, 2025).

We also exploited the model's sensitivity to individual differences in learning to more rigorously test how decision strategies change with age. In line with both previous developmental studies (Decker et al., 2016; Nussenbaum, Scheuplein, et al., 2020; Potter et al., 2017) and our initial hypothesis, we found evidence for developmental changes in MB learning but not MF learning: β_{MB} significantly increased with increasing age ($\beta = 0.04$, $SE = 0.02$, $p = .012$), but β_{MF} did not ($\beta = 0.02$, $SE = 0.01$, $p = .129$). Here, we also found that β_{SR} did not significantly vary with age ($\beta = 0.03$, $SE = 0.02$, $p = .117$), meaning that we did not observe evidence for developmental changes in how people leveraged learned, predictive representations to guide their decisions.

6.4. Flexible arbitration between MB and SR-based learning

Finally, we asked whether participants across age flexibly up- or down-regulated their use of an SR-based decision strategy depending on the predictability of the environment. In stable environments, cached representations of environmental structure are useful. However, if the structure of the environment changes, then relying on a learned transition structure may be maladaptive because such knowledge may rapidly become irrelevant. In changing environments, participants may need to rely on MB learning to a greater extent, because such computations enable greater behavioral flexibility. In this learning task, we did not manipulate the transition structure of the environment directly, but instead manipulated shop reward probabilities, inducing greater change in the transitions that participants experienced in some blocks. Across all blocks, the shop with the highest probability of yielding reward changed islands, imposing continued learning demands on participants throughout the entire task. Importantly, however, the 'predictive representations' leveraged by an SR-based learner depend on their relative probabilities of choosing each of the two shops on each island. Thus, shop reward probabilities changed between blocks in two ways, to either conform with or violate these predictions (Fig. 3B). In *congruent* blocks, the best shop on each island remained the same as they were in the previous block. Here, participants could still rely on their previously learned transition predictions, because they should continue to choose the same shop on each island. In *incongruent* blocks, however, the most rewarding shop on each island also changed. Here, participants' learned predictive representations no longer reflect their likely decisions — if participants learned that they typically visited Shop A on Island 1 but now Shop B is more rewarding, then their more abstract representation of Island 1 may now *over*-incorporate Shop A, which has become largely irrelevant to the overall value of selecting the island. On early trials in incongruent blocks, relying on the previously learned SR will be maladaptive because the participants' shop preferences on each island should change.

To test whether participants flexibly downregulated their use of an SR-based decision strategy on incongruent blocks, we fit a variant of our 'Mixture of Agents' reinforcement learning model that allowed for changes in the contributions of the three learning 'agents' across block types. Following the methods from the prior adult study (Kahn & Daw, 2025), here, rather than directly fitting β_{MF}, β_{MB}, and β_{SR}, we fit two new parameters for each block type: β_{MBSR} and w_{SR}, which enabled us to more directly test our arbitration hypothesis. The β_{MBSR} parameter, defined as $\beta_{MB} + \beta_{SR}$, reflects the *overall* contribution of both the MB and SR-based

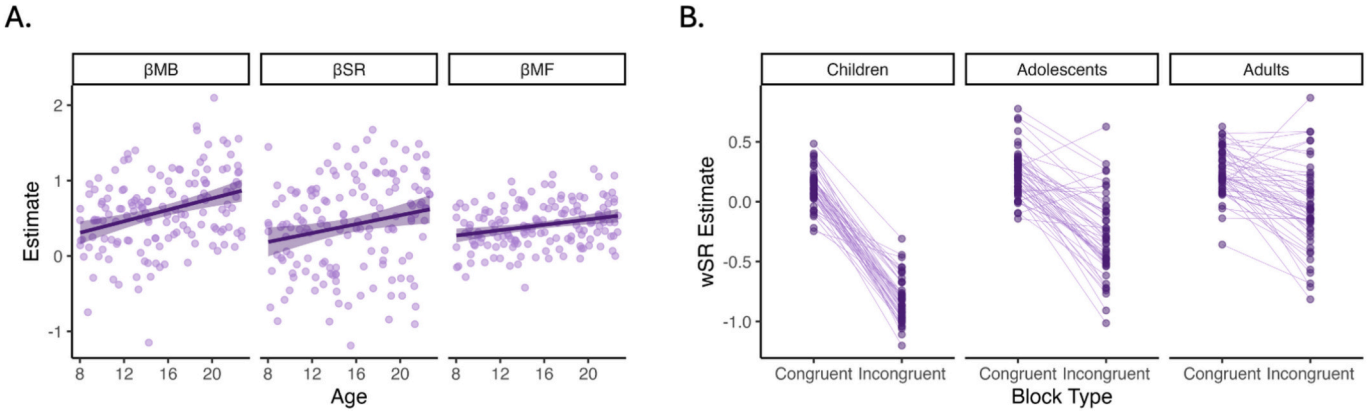


Fig. 5. A) The mixture of agents model revealed that participants increasingly used a MB learning strategy across development. The weight that participants placed on the SR-based and MF learning strategy did not significantly change with age. B) The block-wise mixture of agents model revealed that across age, participants relied on SR-based learning strategy to a greater extent in congruent relative to incongruent blocks. w_{SR} estimates reflect normally distributed parameter values; within the model w_{SR} values were passed through the unit normal cumulative distribution function and transformed to be between 0 and 1. In both panels, points reflect fitted parameter means estimated for individual participants via the hierarchical model. In panel A, the line shows the best-fitting linear regression through the points and the shaded region shows the 95 % confidence interval.

agent, while the w_{SR} parameter, defined as $\frac{\beta_{SR}}{\beta_{MBSR}}$, reflects the *relative* contribution of the SR-based agent. In this way, w_{SR} estimates directly reflect how much participants relied on the SR-based versus MB decision strategy, while accounting for overall, more general individual differences in the use of structured knowledge to guide choice. If participants adapted their decision strategies to the structure of the environment, we would expect w_{SR} in congruent blocks to be higher than w_{SR} in incongruent blocks. All estimated parameters were normally distributed; Within the model, normally distributed w_{SR} values were passed through the unit normal cumulative distribution function and transformed to be between 0 and 1.

In line with the findings from the prior adult study (Kahn & Daw, 2025), participants demonstrated flexible arbitration between decision strategies across block types: w_{SR} in congruent blocks (normally distributed parameter mean = 0.20, $SE = 0.12$) was significantly higher than w_{SR} in incongruent blocks (normally distributed parameter mean = −0.39; $SE = 0.21$; $t(1668) = 5.12$, $p < .001$). Here, we further asked whether 'meta-control' of decision strategies improved with age. Initially, we hypothesized that we might see increasing flexibility across development, such that relative to younger participants, older participants would demonstrate both greater use of the SR in congruent blocks and reduced use of the SR in incongruent blocks. This would be reflected in an age-related increase in w_{SR} in congruent blocks and a decrease in w_{SR} in incongruent blocks. However, we did not observe evidence for age-related changes in w_{SR} in either block type ($ps > 0.07$, see Fig. 5B).

7. Interim discussion

In Experiment 3, we characterized reliance on model-free, model-based and SR-based strategies in a multi-trial reinforcement-learning task with participants aged 8–22 years. In line with past work, we found that use of model-based learning strategies increased with age, while use of model-free learning strategies did not change significantly across development (Cohen et al., 2020; Decker et al., 2016; Nussenbaum, Scheuplein, et al., 2020; Palminteri et al., 2016; Potter et al., 2017; Smid, Ganesan, et al., 2023). Critically, here we also found that participants demonstrated signatures of SR-based learning, which did not change significantly over development.

Additionally, we replicated and extended the prior adult finding that participants adaptively weight their use of MB and SR-based strategies to rely less on the SR when its cached predictive representation is no longer reflective of current transitions (Kahn & Daw, 2025). Despite past work demonstrating developmental differences in meta-control (Bolenz & Eppinger, 2022; Smid, Ganesan, et al., 2023; Smid, Kool, et al., 2023), we found no significant age-related differences in flexible arbitration across our developmental sample. Future work could further investigate how children and adolescents trade off between a wider variety of strategies based on task demands.

8. General discussion

Beyond the well-established dichotomy of model-based and model-free strategies for decision making, there exists a continuum of decision strategies that trade off flexibility and efficiency. In this work, we examined whether children and adolescents make use of "intermediate" learning strategies to support flexible behavior. In Experiments 1 and 2, we asked whether children and adolescents leverage offline replay to flexibly update their behavior when rewards in the environment change. We demonstrated that from childhood to early adulthood, participants used structured task knowledge to guide their choices, but that the opportunity for offline processing during rest did not significantly influence their behavior. In Experiment 3, we showed that like adults, children and adolescents relied on another strategy that balances flexibility with computational efficiency, namely the use of predictive representations.

Though we did not find evidence that behavioral flexibility depended on offline processing during rest in Experiments 1 and 2, our results should not be interpreted as evidence for developmental invariance in the role of offline processing in facilitating value-guided decision making. In other tasks, developmental differences in offline processing may indeed relate to differences observed in the use of structured knowledge for decision making and inference (Cohen et al., 2022; Schlichting et al., 2022; Shing et al., 2019). Further, memory replay during offline processing is thought to depend on interactions between the prefrontal cortex and hippocampus, and the connectivity of these regions exhibits notable developmental change through adolescence (Blankenship et al., 2017; Harvey et al., 2023). Additionally, in rodents, it has been shown that the distance and speed of replayed spatial sequences increases gradually with age over the course of development (Muessig et al., 2019). Thus, while we did not observe evidence for either age-related changes in reward revaluation or for a role of rest-dependent offline processing in facilitating the flexible updating of behavior in our relatively simple, multi-step decision task, developmental changes in both on-task and offline replay may underpin developmental change in value-guided choice in more complex environments.

Rather than relying on offline replay, in Experiments 1 and 2, participants may have relied on predictive representations to update their choices. In Experiment 3, we directly demonstrated that the use of the SR to guide decision making emerges early in development. This early emergence suggests that learning and using predictive representations is a fundamental feature of human cognition that guides behavior from early in life. By caching predictions about upcoming states, the SR enables behavior that is flexible in the face of changing rewards, and computationally less demanding than iterative, model-based simulation. While prior work has proposed a developmental dissociation between learning structured information about a task and using it to guide decision making (Hartley et al., 2021), this dissociation itself may be overly simplified. Here we see that children are able to learn *and* use structured information in the form of the SR. Therefore, it is likely not the case that children learn but do not use structured knowledge, but rather that there are many ways to represent learned structure that can be leveraged to guide adaptive choice to different degrees in different environments. Different tasks may enable or promote reliance on different kinds of knowledge representations (Munakata, 2001), which may explain why some studies find developmental differences in the use of structured knowledge while others do not.

Additionally, while we did not find evidence for developmental changes in the use of the SR, it is possible that there are differences in how the SR is learned and used that did not emerge in our particular task context. Our task consisted of two-stage decisions with only two choices at each stage, and it is possible that developmental differences in the use of the SR would emerge in tasks that require the learning of more complex predictive representations (Nussenbaum et al., 2025). Learning the SR requires tracking the statistics of experience, and iteratively updating beliefs about which states tend to succeed other states. Prior developmental studies of statistical learning have shown that the ability to extract statistical structure from continual experience emerges early in infancy, but continues to change in subtler ways over the course of development (Forest, Schlichting, et al., 2023). From infancy, statistical learning mechanisms underpin our ability to learn language, object categories, and other patterns present in the natural world (Choi et al., 2020; Gómez, 2002; Kirkham et al., 2002; Saffran et al., 1996; Teinonen et al., 2009). However, evidence suggests that learning the statistics of more complex sequences continues to improve with age (Arciuli & Simpson, 2011; Potter et al., 2017; Schlichting et al., 2017). Additionally, the representations formed through statistical learning may shift across development, with younger children demonstrating biases toward learning specific patterns rather than broader, more generalizable ones (Forest, Abolghasem, et al., 2023; Forest, Schlichting, et al., 2023; Pudhiyidath et al., 2020). In line with this literature, future work could investigate developmental differences in learning predictive

representations from experience, as well as in how differences in predictive representations in turn influence decision making.

The hippocampus is thought to play a crucial role in learning the SR (Garvert et al., 2017; Gershman, 2018; Sagiv et al., 2024; Schapiro et al., 2016; Stachenfeld et al., 2017). Unlike the pronounced changes that occur in cortex through adolescence, the hippocampus demonstrates more rapid developmental changes in early childhood (Raznahan et al., 2014; Wierenga et al., 2014), which may facilitate the early learning of predictive representations. Less is known about the neurocognitive mechanisms involved in *using* the SR to guide decision making. Prior work suggests that representations of state predictions in sensory cortex during choice might be involved in use of the SR (Russek et al., 2021) — while here we observed early-emerging use of the SR, it is possible that these cortical representations, or interactions between the hippocampus and cortex (Blankenship et al., 2017; Harvey et al., 2023; Mills et al., 2016; Somerville & Casey, 2010) — may change across development, leading to differences in how the SR is used to guide choice in environments with greater structural complexity.

With many potential decision strategies in their toolkits, children and adolescents still face the challenge of deploying those that are most effective for making decisions in diverse contexts. In this work, we show that participants across age adaptively trade off between using SR-based and MB strategies based on task demands. This suggests that the engagement of different learning and decision strategies is sensitive to environmental structure. Specifically, here we replicated past findings in adults (Kahn & Daw, 2025) demonstrating that the use of the SR emerges most strongly in more stable and predictable environments, where it is most useful. This finding raises the interesting possibility that early experience in predictable environments might facilitate the emergence of this general decision strategy (Birn et al., 2017; Mittal et al., 2015). Future work could investigate this possibility by looking at how the predictability of early life environments influences developmental trajectories of the use of predictive representations in contexts where such representations are both adaptive and maladaptive (Harhen & Bornstein, 2024; McLaughlin et al., 2021; Nussenbaum & Hartley, 2024). Additionally, previous work has shown that people adapt not only their decision strategies, but also the weight they place on recent outcomes in response to reward changes in the environment (Behrens et al., 2007; Kao et al., 2020; Piray & Daw, 2024). There is evidence that even infants and young children adjust how they learn from reward in response to the stability of the environment (Neil et al., 2025; Poli et al., 2025), suggesting early-emerging sensitivity to environmental structure.

Here, across three experiments, we found that when making decisions in environments that allow for the use of computationally efficient learning strategies, children leverage structured knowledge to guide their choices. We further demonstrate that children, adolescents, and adults all make similar use of predictive representations to make adaptive choices. Our findings demonstrate the need for developmental researchers to move beyond simple dichotomies between learning algorithms to take into account how the properties of different learning environments may enable the effective use of different strategies. Grappling with such complexity will deepen our understanding of how experiences in different environments facilitate the emergence and engagement of adaptive strategies for flexible, value-guided decision making across development.

CRediT authorship contribution statement

Alice Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ari E. Kahn:** Writing – review & editing, Software, Methodology, Formal analysis. **Nathaniel D. Daw:** Writing – review & editing, Supervision, Methodology. **Kate Nussenbaum:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Project administration, Methodology, Formal analysis, Conceptualization. **Catherine A. Hartley:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

This work was supported by the National Institute of Mental Health (R01MH126183 to C.A.H., F31MH129105 to K.N., R01MH136875 and R01MH135587 to N.D.) and the CV Starr Foundation Fellowship (to K. N.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106340.

Data availability

All task code, anonymized data, and analysis code is available on the Open Science Framework: osf.io/g83rp.

References

Amso, D., Haas, S., McShane, L., & Badre, D. (2014). Working memory updating and the development of rule-guided behavior. *Cognition, 133*(1), 201–210.
Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science, 14*(3), 464–473.
Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology, 35*(1), 48–69.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3). https://doi.org/10.1016/j.jml.2012.11.001
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.
Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*(9), 1214–1221.
Birn, R. M., Roeber, B. J., & Pollak, S. D. (2017). Early childhood stress exposure, reward pathways, and adult decision making. *Proceedings of the National Academy of Sciences of the United States of America, 114*(51), 13549–13554.
Blankenship, S. L., Redcay, E., Dougherty, L. R., & Riggins, T. (2017). Development of hippocampal functional connectivity during childhood. *Human Brain Mapping, 38*(1), 182–201.
Boddez, Y., Baeyens, F., Hermans, D., & Beckers, T. (2011). The hide-and-seek of retrospective revaluation: recovery from blocking is context dependent in human causal learning. *Journal of Experimental Psychology. Animal Behavior Processes, 37*(2), 230–240.
Bolenz, F., & Eppinger, B. (2022). Valence bias in metacontrol of decision making in adolescents and young adults. *Child Development, 93*(2), e103–e116.
Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science, 31*(9), 1161–1173.
Cohen, A. O., Glover, M. M., Shen, X., Phaneuf, C. V., Avallone, K. N., Davachi, L., & Hartley, C. A. (2022). Reward enhances memory via age-varying online and offline neural mechanisms across development. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 42*(33), 6424–6434.
Cohen, A. O., Nussenbaum, K., Dorfman, H. M., Gershman, S. J., & Hartley, C. A. (2020). The rational use of causal inference to guide reinforcement learning strengthens with age. *npj Science of Learning, 5*, 16.
Collins, A. G. E., & Cockburn, J. (2020). Beyond dichotomies in reinforcement learning. *Nature Reviews. Neuroscience, 21*(10), 576–586.
Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69*(6), 1204–1215.
Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*(12), 1704–1711.
Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation, 5*(4), 613–624.
Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From creatures of habit to goal-directed learners: Tracking the developmental emergence of model-based reinforcement learning. *Psychological Science, 27*(6), 848–858.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology, 49*(1), 60–80.
Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron, 80*(2), 312–325.
Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience, 18*(5), 767–772.
Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology, 22*(6), 1075–1081.
Eldar, E., Lièvre, G., Dayan, P., & Dolan, R. J. (2020). The roles of online and offline replay in planning. *ELife, 9*. https://doi.org/10.7554/eLife.56911
Forest, T. A., Abolghasem, Z., Finn, A. S., & Schlichting, M. L. (2023). Memories of structured input become increasingly distorted across development. *Child Development, 94*(5), e279–e295.
Forest, T. A., Schlichting, M. L., Duncan, K. D., & Finn, A. S. (2023). Changes in statistical learning across development. *Nature Reviews Psychology, 2*(4), 205–219.
Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature, 440*(7084), 680–683.
Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *ELife, 6*. https://doi.org/10.7554/eLife.17086
Gershman, S. J. (2018). The successor representation: Its computational logic and neural substrates. *The Journal of Neuroscience, 38*(33), 7193–7200.
Gershman, S. J., Markman, A. B., & Otto, A. R. (2014). Retrospective revaluation in sequential decision making: a tale of two systems. *Journal of Experimental Psychology. General, 143*(1), 182–194.
Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436.
Harhen, N. C., & Bornstein, A. M. (2024). Interval timing as a computational pathway from early life adversity to affective disorders. *Topics in Cognitive Science, 16*(1), 92–112.
Hartley, C. A., Nussenbaum, K., & Cohen, A. O. (2021). Interactive development of adaptive learning and memory. *Annual Review of Developmental Psychology, 3*(1), 59–85.
Harvey, R. E., Robinson, H. L., Liu, C., Oliva, A., & Fernandez-Ruiz, A. (2023). Hippocampo-cortical circuits for selective memory encoding, routing, and replay. *Neuron, 111*(13), 2076–2090. e9.
Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Computational Biology, 7*(4), Article e1002028.
Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Powers, A., Mehta, N., … Casey, B. J. (2014). Adolescent-specific patterns of behavior and neural activity during social reinforcement learning. *Cognitive, Affective, & Behavioral Neuroscience, 14*(2), 683–697.
Kahn, A. E., & Daw, N. D. (2025). Humans rationally balance detailed and temporally abstract world models. *Communications Psychology, 3*(1), 1–11.
Kao, C.-H., Khambhati, A. N., Bassett, D. S., Nassar, M. R., McGuire, J. T., Gold, J. I., & Kable, J. W. (2020). Functional brain network reconfiguration during learning in a dynamic environment. *Nature Communications, 11*(1), 1682.
Kenward, B., Folke, S., Holmberg, J., Johansson, A., & Gredebäck, G. (2009). Goal directedness and decision making in infants. *Developmental Psychology, 45*(3), 809–819.
Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proceedings of the National Academy of Sciences of the United States of America, 113*(45), 12868–12873.
Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83*(2), B35–B42.
Klossek, U. M. H., Russell, J., & Dickinson, A. (2008). The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. *Journal of Experimental Psychology. General, 137*(1), 39–51.
Kominsky, J. F., Gerstenberg, T., Pelz, M., Sheskin, M., Singmann, H., Schulz, L., & Keil, F. C. (2021). The trajectory of counterfactual simulation in development. *Developmental Psychology, 57*(2), 253–268.
Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological Science, 28*(9), 1321–1333.
Kurth-Nelson, Z., Economides, M., Dolan, R. J., & Dayan, P. (2016). Fast sequences of non-spatial state representations in humans. *Neuron, 91*(1), 194–204.
de Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12.
Lengyel, M., & Dayan, P. (2007). Hippocampal contributions to control: The third way. *Advances in Neural Information Processing Systems, 20*. https://proceedings.neurips.cc/paper/2007/hash/1f4477bad7af3616c1f933a02bfabe4e-Abstract.html.
Liljeholm, M., & Balleine, B. W. (2009). Mediated conditioning versus retrospective revaluation in humans: the influence of physical and functional similarity of cues. *Quarterly Journal of Experimental Psychology (2006), 62*(3), 470–482.
Liu, Y., Dolan, R. J., Higgins, C., Penagos, H., Woolrich, M. W., Ólafsdóttir, H. F., … Behrens, T. E. (2021). Temporally delayed linear modelling (TDLM) measures replay in both animals and humans. *ELife, 10*. https://doi.org/10.7554/eLife.66917
Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D., & Dolan, R. J. (2021). Experience replay is associated with efficient nonlocal learning. *Science, 372*(6544). https://doi.org/10.1126/science.abf1357
Lu, J., Xu, M., Yang, R., & Wang, Z. (2020). Understanding and predicting the memorability of outdoor natural scenes. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society, 29*, 4927–4941.
Luna, B. (2009). Developmental changes in cognitive control through adolescence. *Advances in Child Development and Behavior, 37*, 233–278.
Ma, I., Phaneuf, C., van Opheusden, B., Ma, W. J., & Hartley, C. A. (2022). Distinct developmental trajectories in the cognitive components of complex planning. *PsyArXiv*. https://doi.org/10.31234/osf.io/7fqsr
McLaughlin, K. A., Sheridan, M. A., Humphreys, K. L., Belsky, J., & Ellis, B. J. (2021). The value of dimensional models of early experience: Thinking clearly about concepts and categories. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science, 16*(6). https://doi.org/10.1177/1745691621992346
Mills, K. L., Goddings, A.-L., Herting, M. M., Meuwese, R., Blakemore, S.-J., Crone, E. A., … Tamnes, C. K. (2016). Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. *NeuroImage, 141*, 273–281.
Mittal, C., Griskevicius, V., Simpson, J. A., Sung, S., & Young, E. S. (2015). Cognitive adaptations to stressful environments: When childhood adversity enhances adult executive function. *Journal of Personality and Social Psychology, 109*(4), 604–621.
Momennejad, I. (2020). Learning structures: Predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences, 32*, 155–166.
Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *ELife, 7*. https://doi.org/10.7554/eLife.32548
Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour, 1*(9), 680–692.
Muessig, L., Lasek, M., Varsavsky, I., Cacucci, F., & Wills, T. J. (2019). Coordinated emergence of hippocampal replay and Theta sequences during post-natal development. *Current Biology: CB, 29*(5), 834–840.e4.
Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences, 5*(7), 309–315.
Neil, L., Valton, V., Viding, E., Armbruster-Genc, D., Vuong, V., Packer, K., Sharp, M., Roiser, J., & McCrory, E. (2025). Navigating a varying reward environment in childhood and adolescence. *Scientific Reports, 15*(1), 22715.
Nussenbaum, K., Cohen, A. O., Davis, Z. J., Halpern, D. J., Gureckis, T. M., & Hartley, C. A. (2020). Causal information-seeking strategies change across childhood and adolescence. *Cognitive Science, 44*(9), Article e12888.
Nussenbaum, K., & Hartley, C. A. (2024). Understanding the development of reward learning through the lens of meta-learning. *Nature Reviews Psychology, 3*(6), 424–438.
Nussenbaum, K., Kahn, A., Zhang, A., Daw, N., & Hartley, C. A. (2025). Shifts in learning dynamics drive developmental improvements in the acquisition of structured knowledge. *PsyArXiv*. https://doi.org/10.31234/osf.io/amvth_v1
Nussenbaum, K., Martin, R. E., Maulhardt, S., Yang, Y. J., Bizzell-Hatcher, G., Bhatt, N. S., … Hartley, C. A. (2023). Novelty and uncertainty differentially drive exploration across development. *ELife, 12*. https://doi.org/10.7554/eLife.84260
Nussenbaum, K., Scheuplein, M., Phaneuf, C. V., Evans, M. D., & Hartley, C. A. (2020). Moving developmental research online: Comparing in-lab and web-based studies of model-based reinforcement learning. *Collabra. Psychology, 6*(1), 17213.
Ólafsdóttir, H. F., Carpenter, F., & Barry, C. (2017). Task demands predict a dynamic switch in the content of awake hippocampal replay. *Neuron, 96*(4), 925–935. e6.
Palminteri, S., Kilford, E. J., Coricelli, G., & Blakemore, S.-J. (2016). The computational development of reinforcement learning during adolescence. *PLoS Computational Biology, 12*(6), Article e1004953.
Piray, P., & Daw, N. D. (2021). Linear reinforcement learning in planning, grid fields, and cognitive control. *Nature Communications, 12*(1), 4942.
Piray, P., & Daw, N. D. (2024). Computational processes of simultaneous learning of stochasticity and volatility in humans. *Nature Communications, 15*(1), 9073.
Poli, F., Ghilardi, T., Bersee, J. H. M., Mars, R. B., & Hunnius, S. (2025). Volatility-driven learning in human infants. *Science Advances, 11*(26), Article eadu2014.
Potter, T. C. S., Bryce, N. V., & Hartley, C. A. (2017). Cognitive components underpinning the development of model-based learning. *Developmental Cognitive Neuroscience, 25*, 272–280.
Pudhiyidath, A., Roome, H. E., Coughlin, C., Nguyen, K. V., & Preston, A. R. (2020). Developmental differences in temporal schema acquisition impact reasoning decisions. *Cognitive Neuropsychology, 37*(1–2), 25–45.
R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.
Raab, H. A., & Hartley, C. A. (2018). Chapter 13 - the development of goal-directed decision-making. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-directed decision making* (pp. 279–308). Academic Press.
Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: from childhood to adulthood. *Journal of Experimental Child Psychology, 114*(3), 389–404.
Raznahan, A., Shaw, P. W., Lerch, J. P., Clasen, L. S., Greenstein, D., Berman, R., … Giedd, J. N. (2014). Longitudinal four-dimensional mapping of subcortical anatomy in human development. *Proceedings of the National Academy of Sciences of the United States of America, 111*(4), 1592–1597.
Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Computational Biology, 13*(9), Article e1005768.
Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2021). Neural evidence for the successor representation in choice evaluation. *bioRxiv*. https://doi.org/10.1101/2021.08.29.458114 (p. 2021.08.29.458114).
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.), 274*(5294), 1926–1928.
Sagiv, Y., Akam, T., Witten, I. B., & Daw, N. D. (2024). Prioritizing replay when future goals are unknown. *bioRxivorg*. https://doi.org/10.1101/2024.02.29.582822

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus, 26*(1), 3–8.

Schlichting, M. L., Guarino, K. F., Roome, H. E., & Preston, A. R. (2022). Developmental differences in memory reactivation relate to encoding and inference in the human brain. *Nature Human Behaviour, 6*(3), 415–428.

Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2017). Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of Cognitive Neuroscience, 29*(1), 37–51.

Schuck, N. W., & Niv, Y. (2019). equential replay of nonspatial task states in the human hippocampus. *Science (New York, N.Y.), 364*(6447). https://doi.org/10.1126/science.aaw5181

Shing, Y. L., Finke, C., Hoffmann, M., Pajkert, A., Heekeren, H. R., & Ploner, C. J. (2019). Integrating across memory episodes: Developmental trends. *PLoS One, 14*(4), Article e0215848.

Singer, A. C., Carr, M. F., Karlsson, M. P., & Frank, L. M. (2013). Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. *Neuron, 77*(6), 1163–1173.

Singer, A. C., & Frank, L. M. (2009). Rewarded outcomes enhance reactivation of experience in the hippocampus. *Neuron, 64*(6), 910–921.

Smid, C. R., Ganesan, K., Thompson, A., Cañigueral, R., Veselic, S., Royer, J., … Steinbeis, N. (2023). Neurocognitive basis of model-based decision making and its metacontrol in childhood. *Developmental Cognitive Neuroscience, 62*(101269), Article 101269.

Smid, C. R., Kool, W., Hauser, T. U., & Steinbeis, N. (2023). Computational and behavioral markers of model-based decision making in childhood. *Developmental Science, 26*(2), Article e13295.

Somerville, L. H., & Casey, B. J. (2010). Developmental neurobiology of cognitive control and motivational systems. *Current Opinion in Neurobiology, 20*(2), 236–241.

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar, N., Insel, C., & Wilson, R. C. (2017). Charting the expansion of strategic exploratory behavior during adolescence. *Journal of Experimental Psychology. General, 146*(2), 155–164.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience, 20*(11), 1643–1653.

Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *SIGART Newsletter, 2*(4), 160–163.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience, 10*, 21.

Vikbladh, O. M., Russek, E. M., & Burgess, N. (2024). Consolidation of sequential planning. *bioRxiv*. https://doi.org/10.1101/2024.11.01.621475 (p. 2024.11.01.621475).

Wierenga, L., Langen, M., Ambrosino, S., van Dijk, S., Oranje, B., & Durston, S. (2014). Typical development of basal ganglia, hippocampus, amygdala and cerebellum from age 7 to 24. *NeuroImage, 96*, 67–72.

Wimmer, G. E., Liu, Y., McNamee, D. C., & Dolan, R. J. (2023). Distinct replay signatures for prospective decision-making and memory preservation. *Proceedings of the National Academy of Sciences of the United States of America, 120*(6), Article e2205211120.

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science, 338*(6104), 270–273.