Developmental differences in exploration reveal underlying differences in structure inference

Nora C. Harhen[1], Rheza Budiono[1], Catherine A. Hartley[1*], and Aaron M. Bornstein[2,3*]

[1]Department of Psychology, New York University

[2]Department of Cognitive Sciences, University of California, Irvine

[3]Center for the Neurobiology of Learning & Memory, University of California, Irvine

[*]Denotes equal contribution

Word Count: 4784 words

Author note: All data and code are available at

https://github.com/noraharhen/Harhen-Budiono-Hartley-Bornstein-2025-Foraging

Abstract

Across development, we balance exploring to refine our causal models of the world with exploiting what we already know. Children and adolescents often explore more than adults, a tendency commonly attributed to greater decision-making noise or stronger motivation to learn. Here, we propose developmental changes in structure learning as an alternative driver of exploration. Through exploration, we uncover relevant statistical relationships in our environment that can be leveraged to obtain rewards. To test this proposal, 252 8-to-25-year-old participants completed a patch-foraging task indexing individual differences in structure learning. Younger participants explored more, leaving patches sooner than adults. Computational modeling revealed that their early departures stemmed from their use of simpler and easier-to-plan-over structure representations to guide their decisions. Our findings go beyond previous algorithmic accounts of developmental change in exploration, suggesting that heightened exploration can also arise from differences in how children learn the structure of their environments.

Keywords: exploration, structure learning, foraging, development

Public Significance Statement

Children and adolescents explore more than adults but why? This study suggests that adults recognize more complexity in their environments and strategically adjust their exploration based on subtle differences. In contrast, children and adolescents perceive the world in simpler terms and as a result explore their environments more broadly and consistently.

Study Disclosures

Conflicts of interest: All authors declare no conflicts of interest. Funding: This work was supported by the National Institute of Mental Health (R01 MH126183 to CAH, P50 MH096889 to AMB, PI: TZ Baram, F31 MH134620 to NCH), the Templeton World Charity Foundation (to CAH), the New York University Vulnerable Brain Project (to CAH), and the Department of Defense (NDSEG fellowship to NCH). Ethics: This research received approval from New York University's Institutional Review Board (ID:2021-5210). Preregistration: No aspects of the

study were preregistered. Materials: All study materials are publicly available on github (https://github.com/noraharhen/Harhen-Hartley-Bornstein-2025-Foraging/tree/main/run_exp). Data: All primary data are publicly available on github (https://github.com/noraharhen/Harhen-Hartley-Bornstein-2025-Foraging/tree/main/data_analysis/data) Analysis scripts: All analysis scripts are publicly available on github (https://github.com/noraharhen/Harhen-Hartley-Bornstein-2025-Foraging/tree/main/data_analysis).

Introduction

Human learners face the challenge of discovering the causal structure of their complex and dynamic environments. This problem is particularly daunting for children and adolescents as they have the most to learn and the least prior knowledge to guide their efforts. Fortunately, young learners are adept explorers, actively gathering observations and testing hypotheses to refine their causal inferences (Blanco & Sloutsky, 2021, 2024; Gopnik et al., 2017; Liquin & Gopnik, 2022; Schulz, Wu, Ruggeri, & Meder, 2019; Sumner et al., 2019). Exploration, however, is costly. It often demands time, effort, and forgoing immediate rewards. Balancing the need to gain new information with the benefits of exploiting prior knowledge is a fundamental dilemma for decision makers of all ages. Yet, children and adults approach this trade-off between exploration and exploitation differently. Children and adolescents explore more extensively but less strategically than adults, favoring broad and often stochastic sampling over thorough, targeted exploration (Christakou et al., 2013; Giron et al., 2023; Jepma, Schaaf, Visser, & Huizenga, 2020; Lloyd, McKay, Sebastian, & Balsters, 2021; Meder, Wu, Schulz, & Ruggeri, 2021; Somerville et al., 2017; Wan & Sloutsky, 2024). This developmental shift raises a key question: What mechanisms underlie these changes?

Reinforcement learning provides a valuable computational framework for mechanistically understanding how exploration changes across development (Giron et al., 2023; Meder et al., 2021; Nussenbaum & Hartley, 2019; Nussenbaum et al., 2023). Reinforcement learning models cast age-related differences in exploration as variations in the algorithms that learners use to evaluate options and integrate information. The complexity and computational demands of these algorithms range from basic random sampling to sophisticated evaluations of the costs and benefits of acquiring new information. Despite their differences, these exploration algorithms share a common constraint: their computational cost is tied to the decision maker's representation of the environment (Gershman, Blei, & Niv, 2010; Koenig & Simmons, 1996). An environment with fewer states or contexts ameliorates the demands of an algorithm that is computationally intensive, requiring fewer iterations. Conversely, an environment with many states compounds the demands of an algorithm that is computationally cheap. This implies that a learner could simplify, or abstract, their representation of the environment to circumvent the computational demands of exploration. Using simplified representations may be a particularly useful decision-making strategy for young learners, who may have more constrained working memory capacities (Bunge & Wright, 2007; Cowan, 2014). In this way, developmental shifts in exploration may reflect not only changes in the complexity of decision-making algorithms but also complementary changes in the complexity of mental representations.

Despite their critical role in exploration, the contribution of mental representations to developmental changes in exploration remains poorly understood. This gap is particularly striking given the profound transformations in mental representations that occur in early childhood. During this period, children dramatically reorganize their semantic and conceptual knowledge and revise their intuitive theories of how the world works (Carey, 2009; Gopnik & Meltzoff, 1998; Karmiloff-Smith, 1995; Keil, 1979). This knowledge is built and reconfigured through structure learning processes that involve the extraction of statistical regularities and the inference of causal contingencies. These processes emerge remarkably early in childhood (Saffran & Kirkham, 2018; Gopnik et al., 2004), but continue to develop and strengthen into young adulthood (Finn, Kharitonova, Holtby, & Sheridan, 2019; Forest, Schlichting, Duncan, & Finn, 2023; Schlichting, Guarino, Schapiro, Turk-Browne, & Preston, 2017; Shufaniya & Arnon, 2018). Together, these findings support the notion that age-related changes in mental representations may contribute to changes in exploration.

Here, we examined how structure learning shapes exploration across development. In our patch-foraging task (Harhen & Bornstein, 2023), participants decided whether to stay or leave patches of resources within a richly structured environment. We used computational modeling to quantify the complexity of participants' mental representations and to assess whether their exploration strategies adapted to their uncertainty over these representations. We hypothesized that younger participants would construct simpler representations, an adaptive response to limited experience and cognitive resources. Our model provides a formal mechanistic account of how these simpler representations drive increased exploration.

Methods

Participants

Our final sample consisted of 252 participants, aged 8 to 25 (M = 17.11 years, SD = 5.29, 128 females, 124 males). This sample included 70 children (8.08 - 12.94 years; M = 10.49, 36 females), 68 adolescents (13.07 - 17.94 years; M = 15.47, 35 females), and 114 adults (18 - 25.83 years; M = 22.14, 57 females). This sample size exceeds those of many prior developmental studies investigating value-guided learning and decision making (Cohen, Nussenbaum, Dorfman, Gershman, & Hartley, 2020; Nussenbaum, Scheuplein, Phaneuf, Evans, & Hartley, 2020; Nussenbaum et al., 2023). We selected this sample size to achieve at least 80% power to detect a small effect size (reflecting individual differences in structure learning) taken from the original adult study (ρ=0.16, (Harhen & Bornstein, 2023)) with α=0.05.

All participants reported normal or corrected-to-normal vision and no history of psychiatric or learning disorders. An additional 45 participants completed the study but were excluded based on the following criteria: difficulty understanding task instructions (failing the instruction comprehension check more than twice, n=4), unusually quick responses (mean reaction time < 200 ms, n=12), and extreme strategies ($n = 14$ for mean planet residence time ± 2 *SD* from group mean; n=4 for fully depleting gem mines on more than 75% of visited planets; n=11 for leaving more than 75% of visited planets immediately after the initial dig). Participants received a $10 Amazon gift card for completing the study and had the opportunity to earn an additional performance-based bonus of up to $2.

We recruited participants through the Hartley lab's database, for which we solicited sign-ups via Facebook and Instagram ads, local science fairs and events, and fliers on New York University's campus. Lab researchers verified each participant's age and identity prior to their participation in the online study.

Task

Participants completed a child-friendly variant of a patch foraging task previously used to examine structure learning in adults (Harhen & Bornstein, 2023). Modifications were made to accommodate younger participants, including shortening the task, enhancing the instructions' clarity, and increasing the maximum decision time.

In the task, participants acted as miners collecting "space gems" across various planets (Fig. 1). Bonus payments were tied directly to the total amount of gems collected. Upon landing on a planet, participants dug once, receiving a gem yield sampled from a Normal distribution ($\mu = 100$, $\sigma = 5$). Subsequent trials on the planet involved deciding whether to stay and continue digging, despite increasingly diminishing yields (decreasing according to a planet-type-specific function, described below), or leave for a new planet, incurring a substantial time cost.

If participants chose to stay, a short animation of their avatar digging was shown (3 sec minus the reaction time for that trial), followed by the gem yield (1.5 sec). If they chose to leave, a longer animation of a rocket ship played (10 seconds minus reaction time), followed by an alien welcoming them to the new planet (5.5 sec). Animation durations were linked to the previous trial's reaction time to ensure decision speed did not affect the overall reward

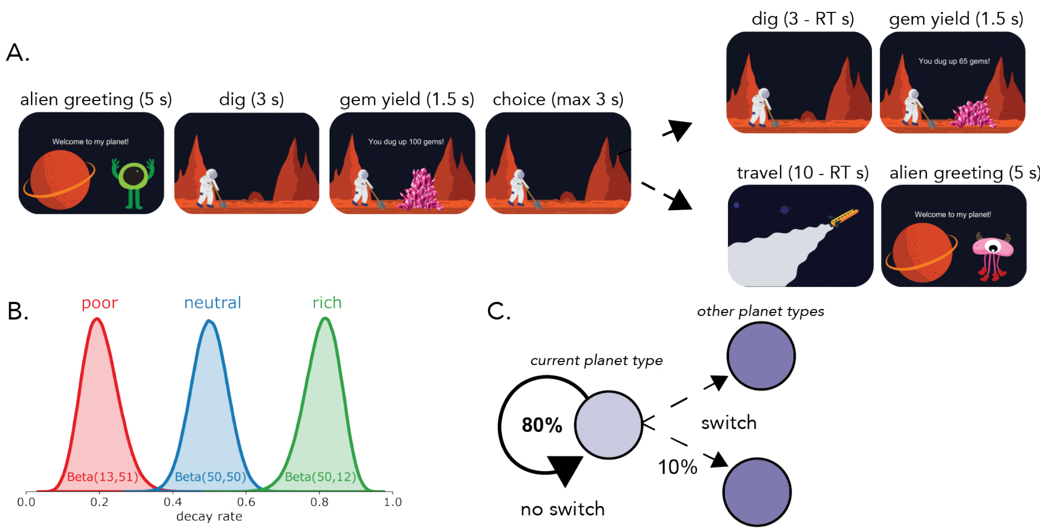

Figure 1: **A. Task design.** Participants traveled to various planets to dig for space gems. On each trial, they decided between continuing to dig on the current planet or traveling to a new one. Both options had their costs: digging depleted the mine, progressively reducing its gem yield, while traveling took a substantial amount of time. **B. Environment structure.** Planets belonged to one of three types–poor, neutral, or rich–differing in how quickly they depleted as mined. Each type was characterized by a distinct distribution over decay rates. **C. Environment dynamics.** Planet richness was correlated in time. A new planet had an 80% probability of being the same type as the previous planet ("no switch") and a 20% probability of transitioning ("switch") to a different type.

rate. Participants had three seconds to respond. If they failed to do so, a red X appeared along with a prompt encouraging them to respond faster. After this, they could re-attempt their choice. In total, the task consisted of four blocks, each lasting six minutes.

The task environment featured three planet types. Rich planets depleted the most slowly (gem yield decreasing with a decay constant sampled from a Beta distribution with parameters α=50, β=12, M=0.8, SD=0.05), poor the most rapidly (α=13, β=51, M=0.2, SD=0.05), and neutral planets at an intermediate rate (α=50, β=50, M=0.5, SD=0.05). On each trial, the depletion rate was newly sampled from the planet's respective distribution. To mimic the structure of natural environments, planet richness was temporally correlated. There was an 80% probability that the next planet would be the same type as the previous one and a 20% chance it would switch to a different type. Importantly, these differences between planets were not explicitly communicated to participants requiring them to infer this information based on the sequence of observed rewards.

We designed the task to measure participants' representational biases independent of task demands. To achieve this, we structured the sequence of planets such that using a more simple representation, one that grouped all planets together, achieved comparable overall rewards to using a more complex representation, one distinguishing between different planet types. Simulations confirmed that the two strategies yielded similar outcomes on average (see supplemental materials S1.1).

Analysis approach

Mixed effects models We used the "lme4" package for R (Bates, Kliegl, Vasishth, & Baayen, 2018) to fit mixed-effects models to our data. Except where noted, models included participant-level random intercepts and random slopes across within-participant fixed effects. To minimize Type I error, we initially specified the maximal model (Barr, Levy, Scheepers, & Tily, 2013). If the model failed to converge, we iteratively simplified the model by first removing interactions between random slopes, followed by random slopes themselves, until convergence was achieved. We used the 'bobyqa' optimizer and set the number of model iterations to 10,000. Continuous variables–age, planet number, and reaction time–were z-scored prior to their inclusion. Age was z-scored across participants while planet number and reaction times were z-scored within. Reaction times were log-transformed before z-scoring.

Marginal Value Theorem To assess the extent to which individuals over- or under-harvested, we compared their planet (patch) resident times to predictions from the Marginal Value Theorem (Charnov, 1976). According to MVT, an optimal agent decides to stay or leave by comparing the immediate expected returns from staying, (V_{stay}), to the opportunity cost of digging on the current planet, (V_{leave}).

V_{stay} is the reward expected from the next dig, the previous reward multiplied by the predicted depletion rate. An MVT-optimal forager accurately identifies the planet's type and uses the true mean of its depletion distribution for their prediction.

$$V_{stay} = r_t * \hat{d} \tag{1}$$

$$\hat{d} = \begin{cases} 0.2 & \text{if planet is poor} \\ 0.5 & \text{if planet is neutral} \\ 0.8 & \text{if planet is rich} \end{cases}$$

Where r_t is the reward received on the last dig, and $\hat{d}$ is the predicted depletion.

They estimate V_{leave}, the expected reward from digging on an alternative planet using the global reward rate, the total rewards received (r_{total}) divided by the total time spent foraging (t_{total}). Reflecting the opportunity cost of digging on the current planet over an unknown alternative planet, they multiply the global reward rate by the time required to dig (t_{dig}).

$$V_{leave} = \frac{r_{total}}{t_{total}} * t_{dig} \tag{2}$$

The forager compares these values and chooses greedily, always selecting the higher-valued action.

Structure learning and uncertainty-adaptive planning model Our model relaxes MVT's assumption of perfect knowledge, introducing two novel computations into the forager's decision-making process: structure learning and uncertainty-adaptive planning.

Foragers must navigate the environment without knowing the true number of planet types, the classification

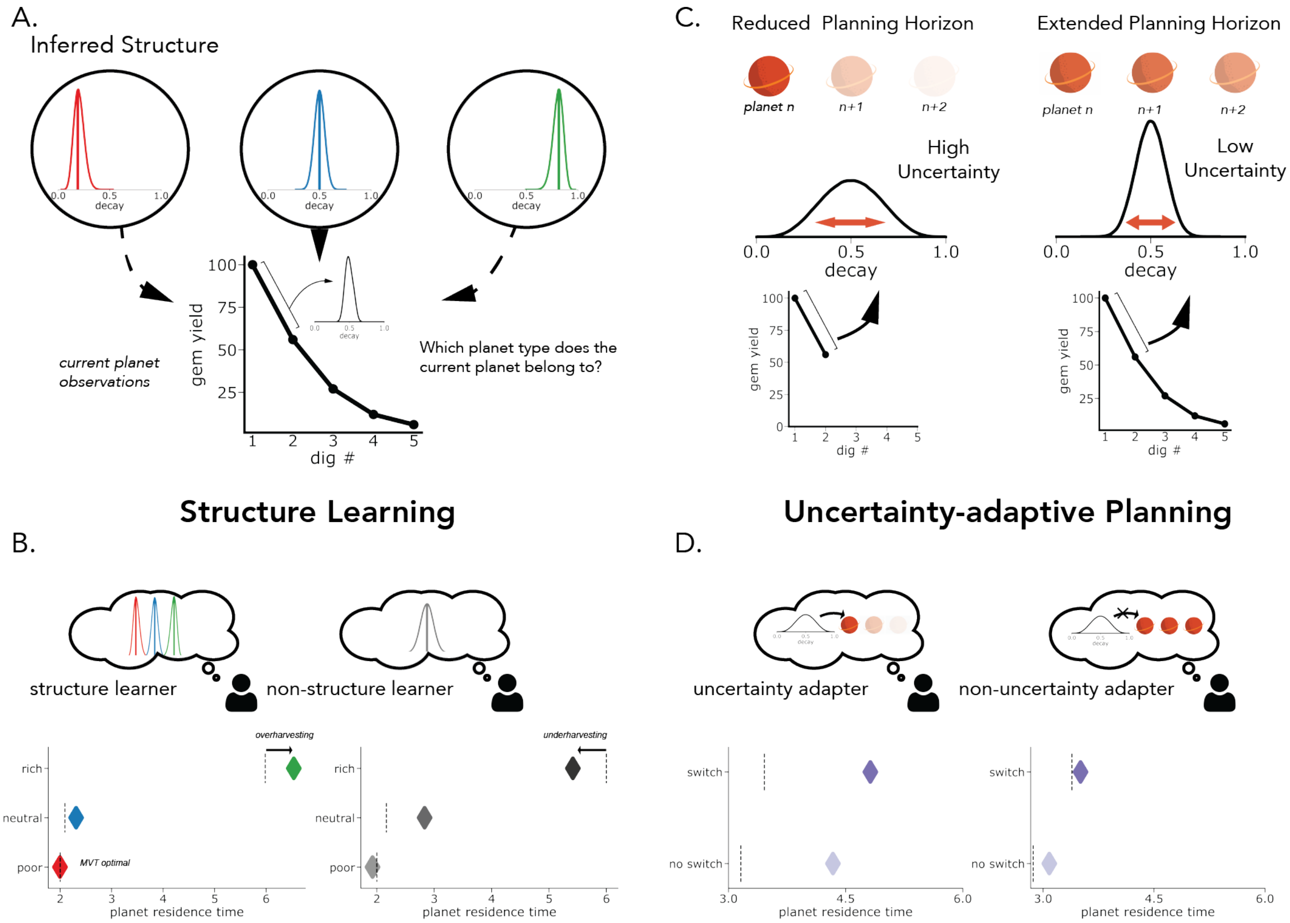

Figure 2: **A. Structure learning computation.** The forager makes two simultaneous inferences: (1) identifying the current planet's type based on observed rewards and (2) determining the total number of planet types in the environment. We model these two inferences as a Chinese Restaurant Process. **B. Structure learning predictions.** The forager's inference of planet types is governed by the parameter, α. The model predicts distinct patterns of over- and under-harvesting depending on the forager's representation of the environment and the number of planet types they consider. The markers show model-simulated planet residence times (PRTs), while dotted lines indicate the Marginal Value Theorem (MVT)-optimal PRTs for reference. **C. Uncertainty adaptive planning computation.** Foragers adjust their planning horizon based on their uncertainty about the current planet's type. Greater uncertainty leads to shorter planning horizons, while lower uncertainty encourages planning further into the future. **D. Uncertainty adaptive planning predictions.** Foragers who adapt their planning horizon to their internal uncertainty should overharvest more following the relatively rare switches in planet type. In contrast, foragers who do not adjust their planning horizon should have more consistent planet residence times across planets.

of individual planets, or the decay rate distribution defining each planet type. To model how foragers infer this information, we use a Chinese Restaurant Process (CRP; (Aldous, 1985)). The CRP's prior is defined around two principles: first, the probability of a planet belonging to an existing type increases with the number of planets already assigned to that type, and second, there remains some probability, proportional to the parameter α, of discovering an entirely new type. This model allows the complexity of foragers' representations to grow as they accumulate experience.

$$P(k) = \begin{cases} \frac{n_k}{N+\alpha} & \text{if k is old} \\ \frac{\alpha}{N+\alpha} & \text{if k is new} \end{cases}$$

Where n_k is the number of planets assigned to type k, α is a clustering parameter, and N is the total number of planets encountered.

After observing a depletion on a planet, the forager computes the posterior probability of the planet being a type as:

$$P(k|D) = \frac{P(D|k)P(k)}{\sum_{j=1}^{J} P(D|j)P(j)} \tag{3}$$

Where J is the number of clusters created up until the current planet, D is a vector of all the depletions observed on the current planet, and all probabilities are conditioned on prior type assignments of planets, $p_{1:N}$

Because computing the exact posterior probability of planet type assignments is computationally intractable, we approximate it using a particle filter with 200 particles. Each particle maintains a hypothetical set of assignments and is weighted based on how well it explains the observed data. When a forager leaves a planet, the particles are resampled. A new pool of particles is generated by sampling with replacement from the previous pool, with the likelihood of a particle being selected proportional to its weight. This process favors particles that better explain the data, increasing their probability of persisting in subsequent iterations.

To predict the next dig's yield, we use a three-step Monte Carlo sampling procedure. First, a particle is selected based on its weight. Second, a planet type is sampled from the selected particle's posterior. Third, a decay rate is drawn from the selected planet type's decay rate distribution. This process is repeated 1,000 times, and the decay rates are averaged over to produce the final prediction.

Each decay rate distribution is initialized to a Gaussian with μ=0.5 and σ=0.5. While the true decay rates follow a Beta distribution, the model assumes normally distributed observations to allow for analytic updates using a Normal-Gamma prior.

Unlike MVT-optimal foragers, who have perfect knowledge, foragers acting according to our model make decisions under epistemic uncertainty. They can never be certain they have an accurate model of the environment. Theoretical work in reinforcement learning suggests that under uncertainty, reducing the planning horizon can improve performance (Jiang, Kulesza, Singh, & Lewis, 2015). In our model, this concept is implemented through an adaptive discount factor, $\gamma_{effective}$, which is defined according to an individual's baseline discounting rate,(γ_{base}), their uncertainty over the current planet's type, (U), and the degree to which uncertainty influences their planning,(γ_{coef}). Uncertainty (U) is measured as the entropy of the Multinomial distribution over planet types, using the same three-step Monte Carlo sampling procedure described above.

$$\gamma_{effective} = \frac{1}{1 + e^{-(\gamma_{base} + \gamma_{coef} * U)}} \tag{4}$$

To model action selection, we use a softmax function, incorporating a lapse rate (ϵ) to account for occasional inattention:

$$p_{stay} = (1 - \epsilon)\frac{1}{1 + e^{-\beta(v_{stay} - v_{leave})}} + \frac{\epsilon}{2} \tag{5}$$

With β being the inverse temperature and ϵ being the lapse rate.

We compared two versions of the model: one with α fixed at 0 and another with α fixed at 0.2. At α=0, the forager assumes that all planets belong to a single type. α=0.2 was the value that in simulation produced the most veridical representation across a range of the model's other parameters (see supplemental materials S1.1). We treated γ_{base}, γ_{coef}, β, and ϵ as free parameters.

Model fitting We fit participants' data on a choice-by-choice basis. Free parameters and their bounds are detailed in the supplementary materials (Table S1). To identify the parameter values that minimized the negative log likelihood of participants' choices, we used Bayesian Adaptive Direct Search (BADS, *Acerbi2017-uu), an optimization algorithm suited for stochastic and computationally expensive functions.

To increase the likelihood of finding the global minimum, we initialized the optimization with different starting points generated from a Sobol sequence. Sobol sequences are quasi-random and have been shown to be more effective than grid or random search (Bergstra & Bengio, 2012), while offering greater computational efficiency than Latin hypercube sampling (Renardy, Joslyn, Millar, & Kirschner, 2021). Starting points were generated until the convergence criteria were met, defined as five consecutive iterations without improvement to the overall minimum. The final parameter values were those that yielded the lowest negative log likelihood across all starting points. Parameter recoverability analyses for both models are included in the supplementary materials (S1.2).

The models' likelihoods are stochastic due to the approximation of the posterior distribution over planet type assignments. To address this noise, we repeated the cluster assignment process 1,000 times. We computed the log likelihood of participants' choices for each of these repetitions and marginalized over them. We then negated this value to obtain the input to the optimization algorithm.

Model comparison We assessed model fit using Akaike Information Criteria (AIC), which penalizes for model complexity. For age-group-level comparisons, we used protected exceedance probabilities (PXP). PXPs estimate the likelihood that a given model is the most frequent best-fitting model within a group while accounting for chance differences in model frequencies. Model recoverability analyses are included in the supplementary materials (S1.2).

Data and code availability statement All data and code are available at `https://github.com/noraharhen/Harhen-Budiono-Hartley-Bornstein-2025-Foraging`.

Results

Model-free

Overharvesting increases with age We first examined how overharvesting varied with age. Using a mixed-effects linear regression model, we assessed the relationship between age and the degree to which planet residence times deviated from Marginal Value Theorem (MVT) optimality. Participants, on average, stayed longer than MVT-optimal (β_0=0.81, $p < .001$), overharvesting planets of their resources. Younger participants, however, tended to leave planets sooner, aligning more closely with MVT optimality (β_{age}=0.22, $p = .045$).

Use of structure knowledge strengthens with age We next tested two key predictions of our model about how overharvesting should vary with planet richness and experience. First, all foragers should overharvest on neutral planets and less so on poor planets, while behavior on rich planets should depend on the forager's representation of the environment (Fig. 2B). Foragers who distinguish planets by richness should overharvest rich planets more than poor ones but less than neutral ones. Neutral planets, falling between these extremes, create the greatest uncertainty and thus encourage the most overharvesting. Conversely, foragers who fail to distinguish planet types should underharvest rich planets. Treating all planets as the same, they overestimate depletion rates for rich planets and leave too soon. Second, our model predicts that as foragers gain more task experience, their representations of the environment should become more accurate, leading them to behave in greater alignment with the Marginal Value Theorem.

To test these predictions, we examined how overharvesting varied with planet richness (i.e., type), task experience (measured by planet number), age, and their interactions. Consistent with the first prediction, participants overharvested the most on neutral planets, the least on poor planets, and to a moderate extent on rich planets (Fig. 3; intercept: β=1.30, $p < .001$; poor planet: β=-0.63, $p < .001$; rich planet: β=-0.42, $p = .0018$). This pattern suggests participants tended to represent the environment as containing multiple distinct planet types. Supporting the second prediction, overharvesting decreased with task experience, particularly on rich planets (planet number: β=-0.24, $p < .001$; planet number x poor planet interaction: β=0.067, $p = .15$; planet number x rich planet interaction: β=-0.26, $p < .001$), indicating that more experience led to more accurate representations of the environment. Overharvesting also varied with age, with the largest differences observed on rich planets (age x rich planet interaction: β=0.36, p = .0078; age: β=0.059, p = .47; age x poor planet interaction: β=-0.045, p = .39). Older participants' behavior suggested greater differentiation between planet types than younger participants. Overall, these results support our model's predictions and demonstrate age-related changes in inferring environmental structure.

Implicit structure knowledge is present across all ages To measure structure learning more implicitly, we assessed reaction times following switches in planet type. Specifically, we focused on participants' second decisions on each planet, as the initial depletion provides the first informative cue about the planet's type. If participants inferred that planets of the same type cluster together in time, they are likely to be surprised–and slower to respond–when a switch in type occurs. To test this prediction, we examined how reaction times changed with switches in planet type, planet number, age, and their interactions using a mixed-effects model, including participant-level random intercepts and slopes for the planet switch regressor. Consistent with our prediction, participants were slower following switches in planet type (switch point: β=0.049, p = .037), suggesting sensitivity to the environment's structure and dynamics. While reaction times generally decreased as the task progressed (planet number: β=-0.049, $p < .001$), post-switch slowing did not significantly diminish with experience (switch x planet number interaction: β=0.014, p = .55). Post-switch slowing did not significantly differ with age (age x switch point interaction: β=0.0088, p = .71), indicating an implicit awareness that planets differed in richness at all ages, even if not evident in choice. These findings suggest a potential developmental dissociation between detecting structure and using it to guide decision making.

Uncertainty adaptive planning emerges early in development We next tested whether participants adjusted their planning horizons based on uncertainty about a planet's type, the key feature of our model's uncertainty-adaptive planning computation. Foragers using this strategy should overharvest more following a switch in planet type, when uncertainty is at its peak (Fig 4). To evaluate this, we examined how overharvesting varied with switches, task experience, age and their interactions. The model included participant-level random intercepts and slopes for planet number. As predicted, participants overharvested more following switches in planet type (switch point: β=0.31, $p < .001$). This effect marginally decreased with experience (switch point $\times$ planet number: β=-0.078, p = .065), consistent with the idea that switches were less uncertainty-inducing as knowledge about the planet structure improved. Early in the task, switch-related overharvesting was similar across ages (age $\times$ switch

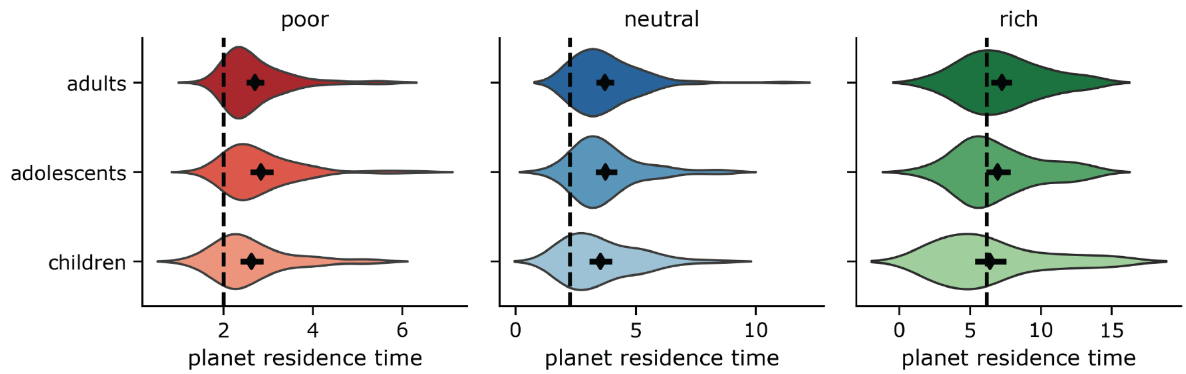

Figure 3: **Model-free signature of structure learning.** Planet residence times (PRT) relative to Marginal Value Theorem (MVT)-optimal PRTs across planet types and age groups. Violin plots show the distribution of PRTs by age group, with markers indicating the means and error bars denoting 95% confidence intervals. Dotted lines provide the MVT-optimal PRT for reference. Overall, participants overharvested across all planet types. Age-related differences were the most pronounced on rich planets. Older participants overharvested on these planets, while many child participants underharvested. Greater variance across child participants on rich planets resulted in a group average PRT that was near MVT-optimal. Our model predictions (Fig 2) suggest that these patterns may reflect age-related differences in the inferred structure of the environment.

point: β=-0.0094, p = .81). However, older participants showed a stronger reduction in switch-related overharvesting with experience (age $\times$ switch point $\times$ planet number: β=-0.11, p = .0082). These findings suggest that while uncertainty-adaptive planning emerges early in development, older participants more readily integrate their knowledge of the environment's structure into their decisions.

Model-based results

Our central question was whether structure learning differs with age. To address this, we fit two models to participants' choices. In one model, α was fixed at 0 and in the other, α was fixed at 0.2, the value that most frequently produced a veridical representation of the environment in simulation (S1.1). Accordingly, we refer to this model as α^* from here on. We compared these models by examining the protected exceedance probabilities (PXP) within each age group (Fig. 5A). Adults' choices were better captured by the α^* model (PXP = 0.88). They learned the environment's structure and used it to guide their decisions. In contrast, children's choices were better captured by the α=0 model (PXP = 0.83). They did not learn to differentiate between the different planet types. Finally, the most frequent, best-fitting model across adolescents was less clear (α^* PXP = 0.66, α=0 PXP= 0.34). To further

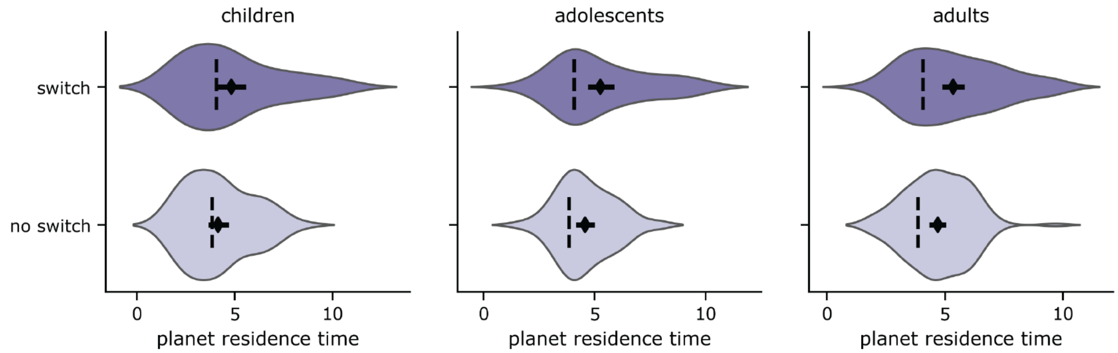

Figure 4: **Model-free signature of uncertainty-adaptive planning.** Planet residence times (PRT) relative to Marginal Value Theorem (MVT)-optimal PRTs across planet types and age groups. Violin plots show the distribution of PRTs by age group, with markers indicating the means and error bars denoting 95% confidence intervals. Dotted lines provide the MVT-optimal PRT for reference. Participants across all age groups adjusted their behavior in response to uncertainty, overharvesting more following rare switches in planet type when uncertainty increased. This behavior is consistent with a reduction in planning horizon.

clarify the developmental trajectory, we examined the relationship between model fit and age, treating the measures as continuous variables. We calculated the difference in AIC scores between the two models. A positive value indicates the participant's choices are better fit by the α^* model. We found that the difference in AIC scores became increasingly positive with age (Fig 5B, Spearman's $\rho = .22$, $p < .001$), supporting a developmental strengthening of structure inference.

Under the α^* model, we found that the baseline discounting factor, the uncertainty-adaptive discounting parameter, the softmax temperature, and the lapse rate did not significantly vary with age (γ_{base}: ρ=-0.049,$p = .44$; γ_{coef}: ρ=-0.068, $p = .28$; β: ρ=0.056, $p = .37$; ϵ: ρ=-0.092, $p = .15$). Fit parameters from the α=0 model yielded the same results (γ_{base} :ρ=-0.041, $p = .52$; β: ρ=0.0095, $p = .88$; ϵ: ρ=-0.087, $p = . = 17$).

Discussion

Here, we investigated structure learning as a potential mechanism underlying children's and adolescents' heightened exploration. In our patch-foraging task, younger participants left planets sooner than adults, with the greatest difference on the richest planets. Despite this, participants of all ages responded similarly to unexpected changes in planet richness, staying longer when the current planet differed in type from the last. Our computational model provided insight into these behaviors. Younger participants left rich planets sooner because they assumed a

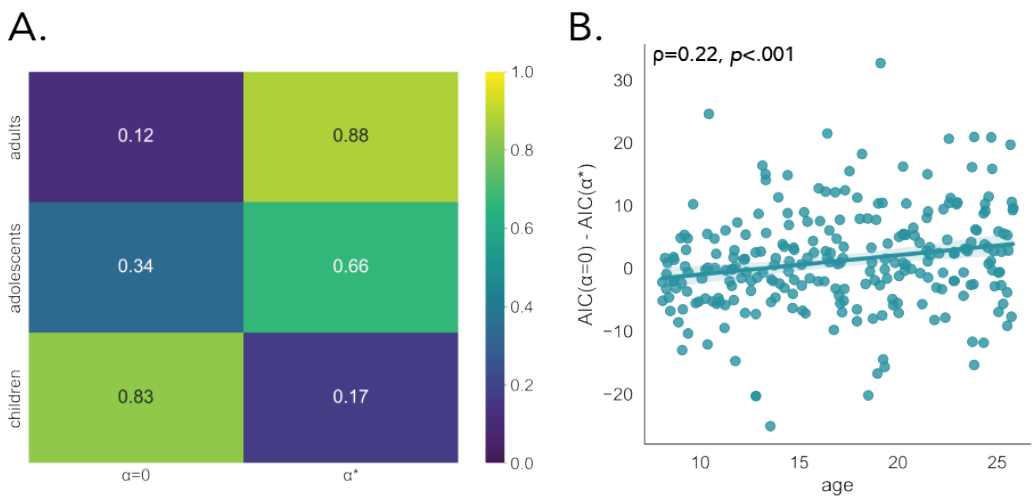

Figure 5: **Model-based results.** **A.** Protected exceedance probabilities (PXPs) reveal differences in the best-fitting model within age groups. The α^* model best captured adults' choices, while the α=0 model better captured children's choices. But, for adolescents, the best-fitting model was less clear. The α^* had a slightly higher PXP than the α=0 model. **B.** We also examined continuous, rather than group-level, age-related differences in model fit. We computed the difference in Akaike Information Criterias (AICs) for each participant and correlated the difference scores with age. A positive difference indicates that the α^* model better captured a participant's choices. We found that the degree to which the α^* model provided a better fit increased with age (Spearman's ρ=0.22, $p < .001$).

simpler environment, grouping all planets into a single type. As a result, they inferred a depletion rate that was faster than the true rate. In contrast, adults stayed longer as they attempted to resolve their uncertainty about the environment's structure, which contained three distinct planet types. Participants' extended stay following a switch in planet type reflects uncertainty-adaptive planning. In other words, they reduced the extent of their planning in response to uncertainty. These findings suggest that developmental differences in exploration stem from how individuals represent the task environment.

A growing body of work suggests that children and adolescents explore more than adults. Two primary explanations have been proposed. One attributes this difference to greater decision-making noise (Giron et al., 2023; Dubois et al., 2022), while the other suggests that children prioritize learning over maximizing immediate rewards (Gopnik, 2020). Our findings complicate both of these accounts. Contrary to the first, computational modeling revealed that children were no noisier than adults. In some ways, their choices were more rational, aligning more closely with the MVT-optimal policy. Moreover, like adults, children deployed sophisticated, flexible planning strategies. In tension with the second explanation, adults' over-exploitation stemmed not from myopic reward-seeking but from efforts to acquire new information about the environment's complex, generative structure. In contrast, children's choices were best explained by a model assuming a simpler representation. These results question whether adults are truly less exploratory than children; perhaps instead, they engage in a different form of exploration that reflects their more nuanced representation of the task. With these results in mind, we propose structure learning as a novel explanation for developmental change in exploration.

Although children behaved as if they used a simpler representation, their reaction times revealed knowledge of a more complex structure. Like adults, they exhibited switch costs, slowing down after encountering a planet that differed in type from the last. Detecting such shifts requires recognizing substantial differences between planets. If children possessed a more accurate representation of the environment, why didn't they use it?

One possibility is that the structural knowledge reflected in children's response times is implicit. Young children often show signs of implicit structural knowledge long before they can explicitly represent or verbally report it (Karmiloff-Smith, 1995). Another possibility is that *using* a complex representation is too cognitively demanding. Past studies have found a developmental dissociation between acquiring structural knowledge and using it to guide behavior (Nussenbaum et al., 2020; Potter, Bryce, & Hartley, 2017; Decker, Otto, Daw, & Hartley, 2016; Cohen et al., 2020). This dissociation may reflect the gradual maturation of the cognitive abilities underpinning the use of mental representations, namely, working memory, future simulation, and episodic memory (Hartley, Nussenbaum, & Cohen, 2021), which continue to mature into young adulthood (Coughlin, Robins, & Ghetti, 2019; Keresztes, Ngo, Lindenberger, Werkle-Bergner, & Newcombe, 2018; Bunge & Wright, 2007). Given these constraints, simpler representations may help children better allocate their limited cognitive resources. From this perspective, the developmental lag between learning and using mental representations may be adaptive. That is, children and adolescents acquire nuanced knowledge of their environments but defer using it until they can do so proficiently.

Alternatively, children may rely on a simpler representation because it is the rational choice, independent of their computational limitations. We designed the task such that using either a simple or complex representation yielded equivalent rewards. Given equal outcomes, the simpler representation becomes the more efficient option, achieving the same goal at a lower computational cost. A better question, then, is why adults rely on the more complex representation when its costs outweigh its benefits. In the absence of incentives for particular representations, they may default to familiar strategies. Their behavior in this task, then, may indicate carryover from their real-world decision-making strategies. Their preference for uncovering causal structure may reflect the demands present in real-world decision-making contexts, where constructing and deploying veridical representations is often beneficial.

Further emphasizing that children are not simply noisier adults, we find that younger participants engaged in uncertainty-adaptive planning, staying longer when the current planet differed in type from the last. At first glance, this seems puzzling. Why would young learners deploy a sophisticated planning strategy when they generally struggle with planning (Nussenbaum et al., 2020; Potter et al., 2017; Decker et al., 2016)? And how could they accomplish this?

Young learners may use uncertainty-adaptive planning because it conserves cognitive resources. One should plan far into the future only when confident in one's own predictions of what will come next. Given children's limited working memory and future simulation abilities, they stand to benefit the most from this computationally frugal approach. Yet, despite these limitations, implementing this strategy may not be particularly demanding. First, it requires reactive not proactive responding. Decision makers adjust in response to surprising outcomes rather than anticipating future ones. This suits children's response tendencies. Early in development, children tend to favor reactive control, which is only engaged when needed, over proactive control, which anticipates control demands in advance (Chevalier, Dauvier, & Blaye, 2018; Chevalier, Meaney, Traut, & Munakata, 2020; Niebaum, Chevalier, Guild, & Munakata, 2021). Second, it requires monitoring internal uncertainty, an early-emerging ability seen in children far younger than those in our study (Baer & Kidd, 2022; Lapidow, Killeen, & Walker, 2022; Schulz et al., 2019; de Eccher, Mundry, & Mani, 2024). These findings challenge the notion that children's exploration is merely a noisier version of adults'. Instead, their strategies are equally sophisticated, sensitive to uncertainty, and well-adapted to their cognitive constraints.

Adults' tendency to stay longer on planets is often interpreted as over-exploitation. However, we argue that, in our task, it can be construed as a form of exploration. Standard definitions describe exploratory actions as ones that will provide information at the cost of immediate reward (Schulz & Gershman, 2019). Staying longer under uncertainty fits this definition. Staying longer allows learners to gather more observations (i.e., depletions) to refine their estimates of the local patch environment and how quickly it will deplete, while also causing them to deviate from the MVT-optimal behavior that would maximize their immediate reward rate. However, in patch-foraging tasks, leaving is typically seen as the exploratory action, as it enables broader environmental sampling and improves estimates of the global reward rate. Staying longer, in contrast, refines knowledge about the local depletion rate. In this sense, both behaviors–leaving sooner and staying longer–serve an exploratory function, but respectively promote broad versus deep exploration of the environment (Moreno-Bote, Ramírez-Ruiz, Drugowitsch, & Hayden, 2020). This reframing suggests that adults' tendency to stay longer may not reflect over-exploitation but rather the use of a deep exploration strategy.

A growing body of work has characterized how exploration changes across development and has primarily proposed age-varying algorithms as the mechanisms underlying these changes. Our findings contribute to this literature by introducing structure learning as a novel mechanism. Rather than being merely noisier or pursuing different objectives than adults, children adaptively adopt representations suited to their computational constraints, allowing them to make decisions and plan in complex environments. More broadly, our findings challenge how we classify behaviors as exploration or exploitation. Individuals may rely on mental representations and information-seeking goals that diverge from those traditionally assumed. When these representations or goals fall outside our hypothesis space, we risk misinterpreting individuals' behavior as noisy, driven by different motivations, or worse, irrational (Acuña & Schrater, 2010; Lapidow & Walker, 2020). Expanding this hypothesis space to include the influence of structured representations invites a richer—and, hopefully, more accurate—understanding of exploration and its developmental trajectory.

References

Acuña, D. E., & Schrater, P. (2010). Structure learning in human sequential decision-making. *PLoS Comput. Biol.*, *6*(12), e1001003.

Aldous, D. J. (1985). Exchangeability and related topics. In D. J. Aldous, I. A. Ibragimov, & J. Jacod (Eds.), *École d'été de probabilités de Saint-Flour XIII — 1983* (pp. 1–198). Springer Berlin Heidelberg.

Baer, C., & Kidd, C. (2022). Learning with certainty in childhood. *Trends Cogn. Sci.*, *26*(10), 887–896.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.*, *68*(3).

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious mixed models.

Bergstra, J., & Bengio, Y. (2012). *Random search for hyper-parameter optimization* (Vol. 13).

Blanco, N. J., & Sloutsky, V. M. (2021). Systematic exploration and uncertainty dominate young children's choices. *Dev. Sci.*, *24*(2), e13026.

Blanco, N. J., & Sloutsky, V. M. (2024). Exploration, exploitation, and development: Developmental shifts in decision-making. *Child Dev.*.

Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Curr. Opin. Neurobiol.*, *17*(2), 243–250.

Carey, S. (2009). *The origin of concepts*. New York, NY: Oxford University Press.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.*, *9*(2), 129–136.

Chevalier, N., Dauvier, B., & Blaye, A. (2018). From prioritizing objects to prioritizing cues: a developmental shift for cognitive control. *Dev. Sci.*, *21*(2).

Chevalier, N., Meaney, J. A., Traut, H. J., & Munakata, Y. (2020). Adaptiveness in proactive control engagement in children and adults. *Dev. Cogn. Neurosci.*, *46*(100870), 100870.

Christakou, A., Gershman, S. J., Niv, Y., Simmons, A., Brammer, M., & Rubia, K. (2013). Neural and psychological maturation of decision-making in adolescence and young adulthood. *J. Cogn. Neurosci.*, *25*(11), 1807–1823.

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 Sci Learn*, *5*, 16.

Coughlin, C., Robins, R. W., & Ghetti, S. (2019). Development of episodic prospection: Factors underlying improvements in middle and late childhood. *Child Dev.*, *90*(4), 1109–1122.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educ. Psychol. Rev.*, *26*(2), 197–223.

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. *Psychol. Sci.*, *27*(6), 848–858.

de Eccher, M., Mundry, R., & Mani, N. (2024). Children's subjective uncertainty-driven sampling behaviour. *R. Soc. Open Sci.*, *11*(4), 231283.

Dubois, M., Bowler, A., Moses-Payne, M. E., Habicht, J., Moran, R., Steinbeis, N., & Hauser, T. U. (2022). Exploration heuristics decrease during youth. *Cogn. Affect. Behav. Neurosci.*, *22*(5), 969–983.

Finn, A. S., Kharitonova, M., Holtby, N., & Sheridan, M. A. (2019). Prefrontal and hippocampal structure predict statistical learning ability in early childhood. *J. Cogn. Neurosci.*, *31*(1), 126–137.

Forest, T. A., Schlichting, M. L., Duncan, K. D., & Finn, A. S. (2023). Changes in statistical learning across development. *Nat. Rev. Psychol.*, *2*(4), 205–219.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychol. Rev.*, *117*(1), 197–209.

Giron, A. P., Ciranka, S., Schulz, E., van den Bos, W., Ruggeri, A., Meder, B., & Wu, C. M. (2023). Developmental changes in exploration resemble stochastic optimization. *Nat Hum Behav*.

Gopnik, A. (2020). Childhood as a solution to explore-exploit tensions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *375*(1803), 20190502.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and bayes nets. *Psychol. Rev.*, *111*(1), 3–32.

Gopnik, A., & Meltzoff, A. N. (1998). *Words, thoughts, and theories.* The MIT Press, Massachusetts Institute of Technology.

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., … Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(30), 7892–7899.

Harhen, N. C., & Bornstein, A. M. (2023). Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. *Proc. Natl. Acad. Sci. U. S. A.*, *120*(13), e2216524120.

Hartley, C. A., Nussenbaum, K., & Cohen, A. O. (2021). Interactive development of adaptive learning and memory.

Annual Review of Developmental Psychology, *3*, 59–85.

Jepma, M., Schaaf, J. V., Visser, I., & Huizenga, H. M. (2020). Uncertainty-driven regulation of learning and exploration in adolescents: A computational account. *PLoS Comput. Biol.*, *16*(9), e1008276.

Jiang, N., Kulesza, A., Singh, S., & Lewis, R. L. (2015). The dependence of effective planning horizon on model accuracy. *Adapt Agent Multi-agent Syst*, 1181–1189.

Karmiloff-Smith, A. (1995). *Beyond modularity.* The MIT Press, Massachusetts Institute of Technology.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective.* Cambridge, MA: Harvard University Press.

Keresztes, A., Ngo, C. T., Lindenberger, U., Werkle-Bergner, M., & Newcombe, N. S. (2018). Hippocampal maturation drives memory from generalization to specificity. *Trends Cogn. Sci.*, *22*(8), 676–686.

Koenig, S., & Simmons, R. G. (1996). The effect of representation and knowledge on goal-directed exploration with reinforcement-learning algorithms. *Mach. Learn.*, *22*(1-3), 227–250.

Lapidow, E., Killeen, I., & Walker, C. M. (2022). Learning to recognize uncertainty vs. recognizing uncertainty to learn: Confidence judgments and exploration decisions in preschoolers. *Dev. Sci.*, *25*(2), e13178.

Lapidow, E., & Walker, C. M. (2020). The search for invariance: Repeated positive testing serves the goals of causal learning. In *Language and concept acquisition from infancy through childhood* (pp. 197–219). Cham: Springer International Publishing.

Liquin, E. G., & Gopnik, A. (2022). Children are more exploratory and learn more than adults in an approach-avoid task. *Cognition*, *218*, 104940.

Lloyd, A., McKay, R., Sebastian, C. L., & Balsters, J. H. (2021). Are adolescents more optimal decision-makers in novel environments? examining the benefits of heightened exploration in a patch foraging paradigm. *Dev. Sci.*, *24*(4), e13075.

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2021). Development of directed and random exploration in children. *Dev. Sci.*, *24*(4), e13095.

Moreno-Bote, R., Ramírez-Ruiz, J., Drugowitsch, J., & Hayden, B. Y. (2020). Heuristics and optimal solutions to the breadth-depth dilemma. *Proc. Natl. Acad. Sci. U. S. A.*, *117*(33), 19799–19808.

Niebaum, J. C., Chevalier, N., Guild, R. M., & Munakata, Y. (2021). Developing adaptive control: Age-related differences in task choices and awareness of proactive and reactive control demands. *Cogn. Affect. Behav.*

Neurosci., *21*(3), 561–572.

Nussenbaum, K., & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Dev. Cogn. Neurosci.*, *40*, 100733.

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*.

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 Psychol.*, *6*(1).

Potter, T. C. S., Bryce, N. V., & Hartley, C. A. (2017). Cognitive components underpinning the development of model-based learning. *Dev. Cogn. Neurosci.*, *25*, 272–280.

Renardy, M., Joslyn, L. R., Millar, J. A., & Kirschner, D. E. (2021). To sobol or not to sobol? the effects of sampling schemes in systems biology applications. *Math. Biosci.*, *337*, 108593.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annu. Rev. Psychol.*, *69*, 181–203.

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. *J. Cogn. Neurosci.*, *29*(1), 37–51.

Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Curr. Opin. Neurobiol.*, *55*, 7–14.

Schulz, E., Wu, C. M., Ruggeri, A., & Meder, B. (2019). Searching for rewards like a child means less generalization and more directed exploration. *Psychol. Sci.*, *30*(11), 1561–1572.

Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cogn. Sci.*, *42*(8), 3100–3115.

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. *J. Exp. Psychol. Gen.*, *146*(2), 155–164.

Sumner, E., Li, A. X., Perfors, A., Hayes, B., Navarro, D., & Sarnecka, B. W. (2019). The exploration advantage: Children's instinct to explore allows them to find information that adults miss. *PsyArxiv*. doi: https://doi.org/10.31234/osf.io/h437v

Wan, Q., & Sloutsky, V. M. (2024). Exploration, distributed attention, and development of category learning. *Psychol. Sci.*, 9567976241258146.