Highlights

Learning Rates Are Not All the Same: The Interpretation of Computational Model Parameters Depends on the Context

Maria K. Eckstein, Sarah L. Master, Liyu Xia, Ronald E. Dahl, Linda Wilbrecht, Anne G.E. Collins

- Many efforts in computational cognitive modeling are based on two implicit assumptions: 1) Model parameters *generalize* between studies and models. 2) Parameters are *interpretable*, i.e., neurally, cognitively, and/or mechanistically specific.
- We empirically investigate these assumptions in Reinforcement Learning (RL), using a large developmental dataset featuring three different learning tasks in a within-participant design.
- We find that RL decision noise/exploration parameters generalize between tasks in terms of captured between-participant variation, but not in terms of absolute parameter values. RL learning rates and Forgetting parameters fail to generalize. All parameters' interpretability is low.
- These findings are in accordance with previous claims that decision noise/exploration decreases during development, but suggest that claims about the development of learning rates or Forgetting should not be generalized between tasks: parameters are unlikely to capture the same neuro-cognitive processes.
- We suggest ways forward to improve the generalizability and interpretability of model parameters, in particular by focusing more on task context.

Learning Rates Are Not All the Same: The Interpretation of Computational Model Parameters Depends on the Context

Maria K. Eckstein[a], Sarah L. Master[a,b], Liyu Xia[a,c], Ronald E. Dahl[a,d], Linda Wilbrecht[a,e], Anne G.E. Collins[a,e]

[a]*Department of Psychology, UC Berkeley, 2121 Berkeley Way West, Berkeley, 94720, CA, USA*
[b]*Department of Psychology, New York University, 6 Washington Place, New York, 10003, NY, USA*
[c]*Department of Mathematics, UC Berkeley, 970 Evans Hall, Berkeley, 94720, CA, USA*
[d]*Institute of Human Development, UC Berkeley, 2121 Berkeley Way West, Berkeley, 94720, CA, USA*
[e]*Helen Wills Neuroscience Institute, UC Berkeley, 175 Li Ka Shing Center, Berkeley, 94720, CA, USA*

Abstract

Reinforcement Learning (RL) models have revolutionized the cognitive and brain sciences, promising to explain behavior from simple conditioning to complex problem solving, to shed light on developmental and individual differences, and to anchor cognitive processes in specific brain mechanisms. However, the growing RL literature shows increasingly contradictory results, potentially casting doubt on the usefulness of the method. We hypothesized that the observed contradictions instead arise from two fundamental assumptions about computational model parameters that are overly strong: That parameters *generalize* between contexts (e.g., tasks, models) and that they capture *interpretable* neurocognitive processes. If these assumptions are corrected, many contradictions dissolve. To test this idea, we asked 291 participants aged 8-30 years to complete three learning tasks during a single experimental session, and fitted their behavior using state-of-the-art RL models. As predicted, we found less-than-perfect parameter generalization between tasks (with better results for RL decision noise/exploration than learning rates) and a fundamental lack of interpretability (especially for learning rates), with a potential role of task similarity for generalization. We conclude that the systematic study of context (including task features and

model parameterization) will enhance the generalizability of computational cognitive models.

Keywords: Computational Modeling, Model Parameter, Reinforcement Learning, Development, Generalizability, Interpretability

1. Introduction

In recent decades, the cognitive neurosciences have made breakthroughs in computational modeling, showing that reinforcement learning (RL) models can explain foundational aspects of human behavior. RL does not only seem to underlie simple cognitive processes such as stimulus-outcome and stimulus-response learning [1, 2, 3], but also highly complex processes, including goal-directed, temporally-extended behavior [4, 5], meta-learning [6], and abstract problem solving that requires hierarchical thinking [7, 8, 9, 10]. Underlining their centrality in the study of human cognition, RL models have been applied across the lifespan [11, 12, 13], and in healthy participants as well those experiencing psychiatric illness [14, 15, 16, 17, 18]. RL models are of particular interest because they promise a close link to brain function: A specialized network of brain regions, including the basal ganglia and prefrontal cortex, implement computations that mirror specific components of RL algorithms, including action values and reward prediction errors [19, 20, 21, 22, 23, 24, 25]. In sum, explaining behaviors from simple conditioning to complex problem solving, adequate for diverse human populations, based on a compelling theoretical foundation [26], and with strong ties to brain function, RL has experienced a surge in published studies since its inception [27], and emerged as a powerful and potentially unifying modeling framework for cognitive and neural processing.

Despite their increasing popularity, however, we do not yet have a sufficient understanding of what exactly RL models and model variables (e.g., model parameters) measure in psychological experiments, and our current assumptions might be imprecise, potentially slowing further progress. Our recent opinion paper develops this argument in depth [28]. In brief, in computational modeling, behavioral datasets are condensed into models and fitted model parameters [11, 27, 29, 30, 31, 32]. These models and parameters are often interpreted as a reflection of ("window into") cognitive or neural processes, with the ability to dissect these into specific, unique components, and to measure participants' inherent characteristics. For example, RL mod-

els have been praised for their ability to dissect choice behavior into value updating and decision making, and to assess individual differences on these components by characterizing individuals in terms of learning rate and decision temperature parameters. However, we argue that these assumptions might be too optimistic and that a careful empirical investigation is required to assess their validity.

To illustrate this point, the current project started out as an investigation into the adolescent development of learning, with the aim of combining the insights of three different learning tasks to gain a more complete understanding of adolescent learning. However, we soon realized that even though each task individually showed strong and interesting developmental patterns [33, 34, 35], these patterns were very different—and even contradictory—across tasks. This showed us that specific model parameters (e.g., learning rate) did not isolate—and measure individuals on—specific cognitive processes (e.g., value updating) that were distinct and theoretically separate from the cognitive processes (e.g., decision making) identified by other parameters (e.g., decision temperature), as we had implicitly assumed. Instead, the "same" parameters seemed to measure something different in each task. This paper explores those questions more formally.

We identified two specific implicit assumptions that had (mis)guided us, which we realized were also adopted widely in the broader computational modeling community [28]: *generalizability* and *interpretability*. We define a model variable (e.g., fitted parameter, reward-prediction error) as *generalizable* if it is consistent across uses, such that a person would be characterized with the same values independent of the specific model or task used to estimate the variable. Generalizability is a consequence of the assumption that parameters are intrinsic to participants rather than task dependent (e.g., a high learning rate is a personal characteristic that might reflect an individual's unique brain structure). We further define a model variable as *interpretable* if it isolates specific and unique cognitive elements, and/or is implemented in separable and unique neural substrates. Interpretability follows from the assumption that the decomposition of behavior into model parameters "carves cognition at its joints" (e.g., separating value updating from decision making).

Assumptions about generalizability and interpretability are rarely stated explicitly, but implicitly underlie work in fields spanning from computational psychology to neuroscience, and shape the way research is planned, conducted, and interpreted. For example, attempting to identify humans'

inherent parameter settings (e.g., empirical parameter distributions [36]; interplay between different learning rates [37]) is based on the assumption of generalizability because parameter are taken to be independent of certain task features but specific to a person. Similarly, the effort to characterize the age development of parameters often assumes that parameters are study-independent but person-specific [11, 12, 13, 33]. And perhaps most generally, comparing model variables (e.g., parameters) between studies in review articles [14, 15, 16, 19, 20, 21, 22, 23, 25], meta-analyses [24, 38, 39], and discussion sections of empirical papers implicitly equalizes variables between studies, thereby assuming generalization. Interpretability, on the other hand, underlies efforts to tie specific model variables to specific neural substrates (e.g., reward prediction error to dopamine function [40]), to use them as markers of psychiatric conditions (e.g., working-memory parameter differences in schizophrenia [41]) or real-world decision making [42], and generally the assumption that model variables are particularly "theoretically meaningful" [14] compared to other, e.g., behavioral, measures.

However, emerging inconsistencies are increasingly becoming the focus of attention in cognitive [43], developmental [13, 44, 45, 46], clinical [15, 16, 17, 18], and neuroscientific research [24, 38, 39, 47]. For example, some developmental studies have shown that learning rates increased with age [33, 48], whereas others have shown that they decrease [49]. Yet others have reported U-shaped trajectories [34], or stability within a given age range [50] (for a comprehensive review, see [13]; for specific examples, see [43]). Our opinion paper elaborates on a range of similar inconsistencies [34]. We suggest that these apparent inconsistencies arise because we mistakenly assume parameter generalizability and interpretability, rather than reflecting deeper problems with the method of computational modeling. This view is supported by several theoretical considerations regarding model fitting [27, 51, 52, 53, 54], and in line with the observation that parameters can be highly context-dependent [13, 43, 38, 55, 56].

The goal of the current project was to assess this hypothesis empirically, determining the degree to which parameters generalize between a set of relatively similar tasks, and investigating whether parameters were interpretable within this context. We compared the RL parameters fit to the same individuals across different learning tasks in a single study, using a developmental dataset (291 participants, ages 8-30 years), which allowed us to obtain a large spread of individual differences and address outstanding discrepancies in the developmental psychology literature [13]. The three learning tasks

varied on several common dimensions, including feedback stochasticity, task volatility, and memory demands (Fig. 1B), and have been used previously to study RL processes [33, 34, 35] (However, like many tasks in the literature, these tasks likely also engaged other cognitive processes besides RL, such as working memory and reasoning.) The within-participant design allowed us to test directly whether the same participants showed the same parameters across tasks (generalizability), and the combination of multiple tasks shed light on which cognitive processes the same parameters captured in each task (interpretability). We extensively compared and validated all RL models [27, 30, 57] and reported each task's unique developmental results separately [33, 34, 35].

Our results show a striking lack of generalizability and interpretability, suggesting that past computational findings are not as generalizable and interpretable as often assumed, and that future research needs to address the reasons of the observed discrepancies to move the field forward.

2. Results

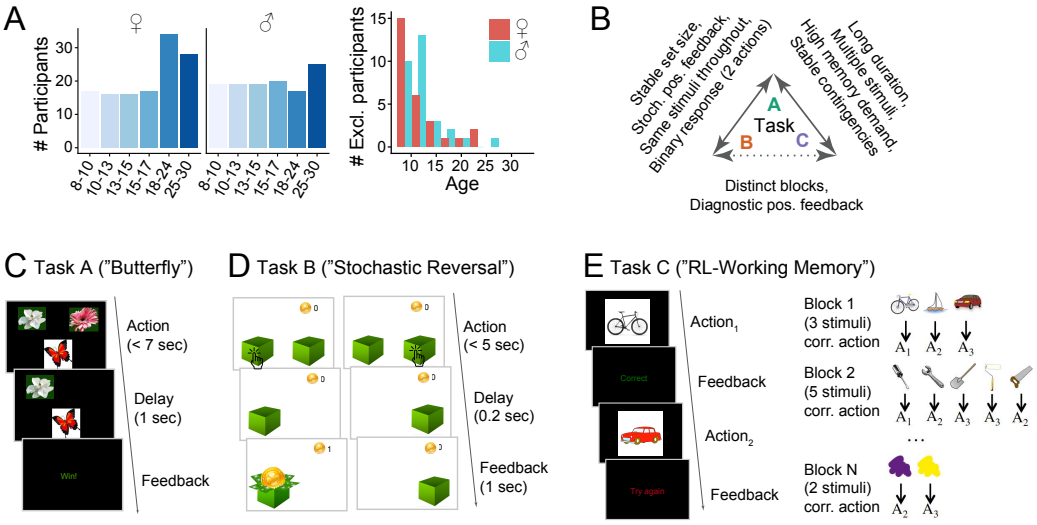

Figure 1: **Overview of the experimental paradigm.** (A) Participant sample. Left: Number of participants in each age group, broken up by sex (self-reported). Age groups were determined by within-sex age quartiles for participants between 8-17 years (see [34] for details) and 5-year bins for adults. Right: Number of participants whose data were excluded because they failed to reach performance criteria in at least one task. (B) Pairwise similarities in terms of experimental design between tasks A [35], B [34], and C [33]. Similarities are shown on the arrows connecting two tasks; the lack of a feature implies a difference. E.g., a "Stable set size" on tasks A and B implies an unstable set size in task C. Overall, task A shared more similarities with tasks B and C than these shared with each other. (C) Task A procedure of ("Butterfly task"). Participants saw one of four butterflies on each trial and selected one of two flowers in response, via button press on a game controller. Each butterfly had a stable preference for one flower throughout the task, but rewards were delivered stochastically (70% for correct responses, 30% for incorrect). For details, see section 4.4 and the original publication [35]. (D) Task B Procedure ("Stochastic Reversal"). Participants saw two boxes on each trial and selected one with the goal of finding gold coins. At each point in time, one box was correct and had a high (75%) probability of delivering a coin, whereas the other was incorrect (0%). At unpredictable intervals, the correct box switched sides. For details, see section 4.4 and [34]. (E) Task C procedure ("Reinforcement learning-working memory"). Participants saw one stimulus on each trial and selected one of three buttons ($A_1 - A_3$) in response. All correct and no incorrect responses were rewarded. The task contained blocks of 2-5 stimuli, determining its "set size". The task was designed to disentangle set-size sensitive working memory processes from set-size insensitive RL processes. For details, see section 4.4 and [33].

This section gives a brief overview of the experimental tasks and computational models (details in Fig. 1C-E, sections 4.4 and 4.5, and original publications [33, 34, 35]), before showing the results on parameter generalizability (section 2.1) and interpretability (section 2.2).

All three tasks considered are reinforcement learning tasks commonly modeled with reinforcement learning models. Depending on the task, RL models contained both shared and different parameters, as is common practice in the literature in order to capture behavior in each task as neatly as possible. Task A required participants to learn the correct associations between each of four stimuli (butterflies) and two responses (flowers) based on probabilistic feedback (Fig. 1C). The best-fitting model contained three free parameters: learning rate from positive outcomes α_+, inverse decision temperature β, and Forgetting F. It also contained one fixed parameter: learning rate from negative outcomes $\alpha_- = 0$ [35]. Task B required participants to adapt to unexpected switches in the action-outcome contingencies of a simple bandit task (only one of two boxes contained a gold coin at any time) based on semi-probabilistic feedback (Fig. 1D). The best-fitting RL model contained four free parameters: α_+, α_-, β, and choice persistence p [34]. Task C required learning of stimulus-response associations like task A, but over several task blocks with varying numbers of stimuli, while using deterministic feedback (Fig. 1E). The best model for this task combined RL and working-memory mechanisms, containing RL parameters α_+ and α_-; working-memory parameters capacity K, Forgetting F, and noise ϵ; and mixture parameter ρ, which determined the relative weights of RL and working memory [33, 58].

To ensure that potential parameter discrepancies in this study were not due to a lack of modeling quality, we employed rigorous model fitting, comparison, and validation [27, 29, 30, 57]: For each task, we compared a large number of competing models, based on different parameterizations and cognitive mechanisms, and selected the best one based on quantitative model comparison scores as well as models' abilities to reproduce participants' behavior in simulation. We used hierarchical Bayesian methods for model fitting and comparison where possible to obtain the most accurate parameter estimates [57].

2.1. Part I: Parameter Generalizability

To investigate parameter generalizability, we first assessed whether participants showed similar parameter values across tasks and then determined

whether tasks showed similar parameter age trajectories. This within-participant comparison is crucial to determine whether the inconsistencies in the previous literature were caused by differences between studies (e.g., between participant samples, testing procedures, modeling approaches, research labs) or reflect a more fundamental lack of generalizability, as we predicted.

2.1.1. Differences in Absolute Parameter Values

We first used repeated-measures analyses of variance (ANOVAs) to test for task effects on absolute parameter values (Fig. 2A). When ANOVAs showed significant task effects, we followed up with repeated-measures t-tests to compare each pair of tasks, using the Bonferroni correction.

Learning rates α_+ and α_- were so dissimilar across tasks that they occupied largely separate ranges: They were very low in tasks C (α_+ mean: 0.07, sd: 0.18; α_- mean: 0.03, sd: 0.13), intermediate in task A (α_+ mean: 0.22, sd: 0.09; α_- was fixed at 0), but fairly high in task B (α_+ mean: 0.77, sd: 0.11; α_- mean: 0.62, sd: 0.14; for statistical comparison, see Table 1). Decision noise was high in task B ($\frac{1}{\beta}$ mean: 0.33, sd: 0.15), but low in tasks A ($\frac{1}{\beta}$ mean: 0.095, sd: 0.0087) and C (ϵ mean: 0.025, sd: 0.032; statistics in Table 1 ignore ϵ because its absolute values were not comparable to $\frac{1}{\beta}$ due to the different parameterization; see section 4.5). Forgetting was significantly higher in task C (mean: 0.19, sd: 0.17) than A (mean: 0.056, sd: 0.028; task B was best fit without forgetting).

For all parameters, absolute parameter values differed substantially between tasks. This shows that the three tasks produced significantly different estimates of learning rate, decision noise/exploration, and forgetting for the same participants (Fig. 2B). Interestingly, these parameter differences echoed specific task demands: Learning rates and noise/exploration were highest in task B, where frequent switches required quick updating and high levels of exploration. Similarly, forgetting was highest in task C, which posed the largest memory demands. Using regression models that controlled for age (instead of ANOVA) led to similar results (Table D.9).

Table 1: Statistics of ANOVAs predicting raw parameter values from task (A, B, C). When an ANOVA showed a significant task effect, we followed up with post-hoc, Bonferroni-corrected t-tests. * $p < .05$; ** $p < .01$; *** $p < .001$.

Parameter	Model	Tasks	F / t	df	p	sig.
$\frac{1}{\beta}$	ANOVA	A, B	830	1	$p < 0.001$	***
	t-test	A vs B	25	246	$p < 0.001$	***
α_+	ANOVA	A, B, C	2.018	2	$p < 0.001$	***
	t-test	A vs B	66	246	$p < 0.001$	***
	t-test	A vs C	12	246	$p < 0.001$	***
	t-test	B vs C	51	246	$p < 0.001$	***
α_-	ANOVA	B, C	2.357	1	$p < 0.001$	***
	t-test	B vs C	49	246	$p < 0.001$	***
Forgetting	ANOVA	A, C	161	1	$p < 0.001$	***
	t-test	A vs C	49	246	$p < 0.001$	***

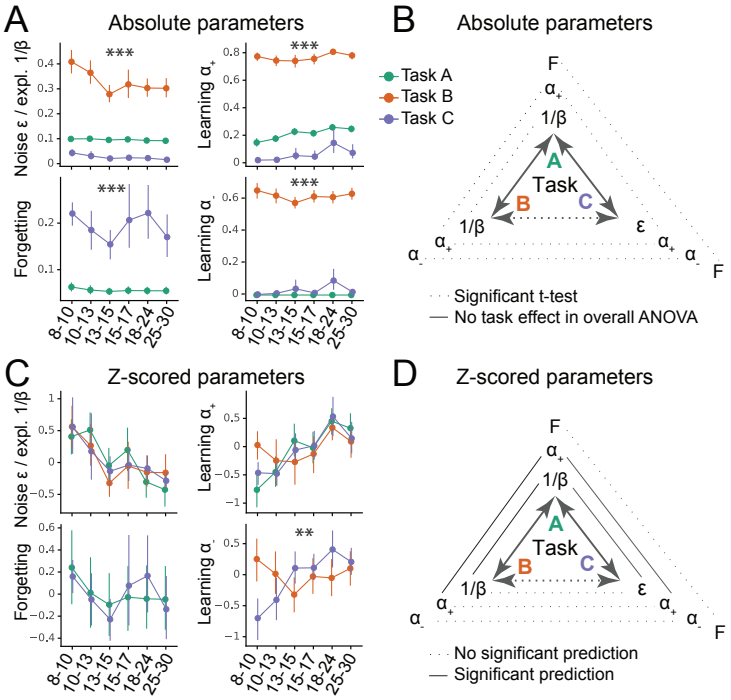

Figure 2: **Generalizability of absolute parameter values (A-B) and of parameter age trajectories / z-scored parameters (C-D).** (A) Fitted parameters over participant age (binned) for all three tasks (A: green; B: orange; C: blue). Parameter values differed significantly between tasks; significance stars show the p-values of the main effects of task on parameters (Table 1; * $p < .05$; ** $p < .01$; *** $p < .001$). Dots indicate means, error bars specify the confidence level (0-1) for interval estimation of the population mean. (B) Summary of the main results of part A. Double-sided arrows connecting tasks are replicated from Fig. 1B and indicate task similarity (dotted arrow: small similarity; full arrow: large similarity). Lines connecting parameters between tasks show test statistics (Table 1). Dotted lines indicate significant task differences in Bonferroni-corrected pairwise t-tests (full lines would indicate the lack of difference). All t-tests were significant, indicating that absolute parameter values differed between each pair of tasks. (C) Parameter age trajectories, i.e., within-task z-scored parameters over age. Age trajectories reveal similarities that are obscured by differences in means or variances in absolute values (part A). Significance stars show significant effects of task on age trajectories (Table 2). (D) Summary of the main results of part C. Lines connecting parameters between tasks show statistics of regression models predicting each parameter from the corresponding parameter in a different task (Table 4). Full lines indicate significant predictability and dotted lines indicate a lack thereof. In contrast to absolute parameter values, age trajectories were predictive in several cases, especially for more similar tasks with more similarities (A and B; A and C), compared to tasks with fewer (B and C).

2.1.2. Relative Parameter Differences

However, comparing parameters in terms of their absolute values has shortcomings because it ignores variance between participants. Nevertheless, reflecting participants' mutual relationships to each other, variance might be an important component of what parameters measure. To test whether this aspect generalized between tasks, we first correlated corresponding parameters between each pair of tasks, using Spearman correlation (suppl. Fig. D.8). Indeed, both α_+ (suppl. Fig. D.8A) and noise/exploration parameters (suppl. Fig. D.8B) were significantly positively correlated between tasks A and B as well as between tasks A and C. Significant correlations were lacking between tasks B and C. This suggests that both α_+ and noise/exploration generalized in terms of the relationships they captured between participants; however, this generalization was only evident between tasks A and B or A and C, potentially due to the fact that task A was more similar to tasks B and C than these were to each other (Fig. 1B; also see section 2.2.1). Fig. D.9 shows the correlations between all pairs of features in the dataset (model parameters and behavioral measures). Note that noise parameters generalized between tasks A and C despite differences in parameterization (ϵ vs. $1/\beta$), showing robustness in the characterization of choice stochasticity (suppl. Fig. D.8B).

2.1.3. Parameter Age Trajectories

This correlation analysis, however, is limited in its failure to account for age, an evident source of variance in our dataset. This means that apparent parameter generalization could be driven by a common dependence on age, rather than underlying age-independent similarities. To address this, we next focused on parameter age trajectories, aiming to remove arbitrary differences between tasks (e.g., absolute mean and amount of variances), while conserving meaningful patterns (e.g., shape of variance, i.e., participants' values relative to each other). Age trajectories were calculated by z-scoring each parameter within each task (Fig. 2C). To test for differences, mixed-effects regression was used to predict parameters of all tasks from two age predictors (age and squared age) and task (A, B, or C). A better fit of this model compared to the corresponding one without task indicates that task characteristics affected age trajectories. In this case, we followed up with post-hoc models comparing individual pairs of tasks.

For α_-, the task-based regression model showed a significantly better fit, revealing an effect of task on α_-'s age trajectory (Table 2). Indeed, α_-

showed fundamentally different trajectories in task B compared to C (in task A, α_- was fixed): In task B, α_- decreased linearly, modulated by a U-shaped curvature (linear effect of age: $\beta = -0.11$, $p < 0.001$; quadratic: $\beta = 0.003$, $p < 0.001$), but in task C, it increased linearly, modulated by an inverse-U curvature (linear: $\beta = 0.32$, $p < 0.001$; quadratic: $\beta = -0.07$, $p < 0.001$; Fig. 2C). The fact that these patterns are opposites of each other was reflected in the significant interaction terms of the grand regression model (Table 3). Indeed, we previously reported a U-shaped trajectory of α_- in task B, showing a minimum around age 13-15 [34], but a consistent increase up to early adulthood in task C [35]. This shows striking difference when estimating α_- using task B compared to C. These differences might reflect differences in task demands: Negative feedback was diagnostic in task C, requiring large learning rates from negative feedback α_- for optimal performance, whereas negative feedback was not diagnostic in task B, requiring small α_-.

For α_+, adding task as a predictor did not improve model fit, suggesting that α_+ showed similar age trajectories across tasks (Table 2). Indeed, α_+ showed a linear increase that tapered off with age in all tasks (linear increase: task A: $\beta = 0.33$, $p < 0.001$; task B: $\beta = 0.052$, $p < 0.001$; task C: $\beta = 0.28$, $p < 0.001$; quadratic modulation: task A: $\beta = -0.007$, $p < 0.001$; task B: $\beta = -0.001$, $p < 0.001$; task C: $\beta = -0.006$, $p < 0.001$). For noise/exploration and Forgetting parameters, adding task as a predictor also did not improve model fit (Table 2), suggesting similar age trajectories across tasks. For decision noise/exploration, the grand model revealed a linear decrease and tapering off with age (Fig. 2C; Table 3), in accordance with previous findings [13]. For Forgetting, however, the grand model did not reveal any age effects (Fig. 2C; Table 3), suggesting inconsistent or lacking developmental changes. In summary, the age trajectories of noise/exploration parameters, α_+, and Forgetting did not differ between tasks, whereas the trajectory of α_- differed strikingly.

2.1.4. Predicting Age Trajectories

The previous analysis, focusing on parameter differences, revealed some *lack* of generalization (e.g., α_-). The next analysis takes the inverse approach, assessing similarities in an effort to provide evidence *for* generalization: We used linear regression to predict participants' parameters in one task from the corresponding parameter on another task, controlling for age and squared age.

For both α_+ and noise/exploration parameters, task A predicted tasks

Table 2: Assessing task effects on parameter age trajectories: Model fits (AIC scores) of regression models predicting parameter age trajectories, comparing the added value of including ("AIC with task") versus excluding ("AIC without task") task as a predictor. Differences in AIC scores were tested statistically using F-tests. Better (smaller) model fits are highlighted in bold. The coefficients of the winning models (simpler model "without task" unless adding task predictor leads to significantly better model fit) are shown in Table 3.

Parameter	AIC without task	AIC with task	F(df)	p	sig.
$\frac{1}{\beta}/\epsilon$	**2,044**	2,054	NA	NA	–
α_+	2,044	**2,042**	$F(4, 245) = 2.34$	$p = 0.056$	–
α_-	1,395	**1,373**	$F(2, 245) = 6.99$	$p = 0.0011$	**
Forgetting	**1,406**	1,411	NA	NA	–

Table 3: Statistical tests on age trajectories: mixed-effects regression models predicting z-scored parameter values from task (A, B, C), age, and squared age (months). When the task-less model fitted best, the coefficients of this model are shown, showing shared age trajectories (Table 2; $\frac{1}{\beta}/\epsilon$, α_+, Forgetting). When the age-based model fitted better, pairwise follow-up models are shown (α_-), showing task differences. P-values of follow-up models were corrected for multiple comparison using the Bonferroni correction. * $p < .05$; ** $p < .01$, *** $p < .001$.

Parameter	Tasks	Predictor	β	p (Bonf.)	sig.
$\frac{1}{\beta}/\epsilon$	A, B, C	Intercept	1.86	< 0.001	***
		Age (linear)	-0.17	0.003	**
		Age (quadratic)	0.004	< 0.001	***
α_+	A, B, C	Intercept	-2.10	< 0.001	***
		Age (linear)	0.20	< 0.001	***
		Age (quadratic)	-0.004	< 0.001	***
α_-	B, C	Task (main effect)	4.15	< 0.001	***
		Task * linear age (interaction)	0.43	< 0.001	***
		Task * quadratic age (interaction)	-0.010	< 0.001	***
Forgetting	A, C	Intercept	0.37	0.44	
		Age (linear)	-0.034	0.53	
		Age (quadratic)	0.001	0.63	

B and C, and tasks B and C predicted task A, but tasks B and C did not predict each other (Table 4; Fig. 2D), reminiscent of the correlation results (section 2.1.2). For α_-, tasks B and C showed a marginally significant *negative* relationship (Table 4), suggesting that predicting α_- between tasks can lead to systematically biased predictions, confirming the striking differences observed before (section 2.1.3). For Forgetting, tasks A and C were not predictive of each other (Table 4), suggesting that the lack of significant differences we observed previously (Table 3) did not necessarily imply successful generalization, but might have been caused by other factors, e.g., elevated noise.

Table 4: Statistics of the regression models predicting each parameter from the corresponding parameter in a different task, while controlling for age. Results were identical when predicting task A from B and task B from A, for all pairs of tasks. Therefore, only one set of results is shown, and predictor and outcome task are not differentiated.

Parameter	Tasks	β	p	sig.
$\frac{1}{\beta}$, ϵ	A & B	0.28	< 0.001	***
	A & C	0.19	0.0022	**
	B & C	0.039	0.54	
α_+	A & B	0.13	0.035	*
	A & C	0.23	< 0.001	***
	B & C	-0.073	0.25	
α_-	B & C	-0.12	0.058	$
Forgetting	A & C	0.097	0.13	

2.1.5. Summary Part I: Generalizability

Summarizing our results so far, no parameter showed generalization between tasks in terms of *absolute values* (Fig. 2A, 2B), but noise/exploration and α_+ showed similar *age trajectories* (Fig. 2C), at least in tasks that were sufficiently similar (Fig. 2D). More specifically, (1) all parameters differed significantly between tasks in terms of absolute values (Fig. 2A, 2B). Intriguingly, absolute parameter values varied more between tasks than between participants within tasks, suggesting that task demands played a larger role in determining parameter values than participants' individual characteristics. This was the case for all four model parameters (Noise/Exploration, α_+, α_-, and Forgetting). (2) However, there was evidence that in some cases, parameter age trajectories generalized between tasks: Task identity did not affect

the age trajectories of noise/exploration, Forgetting, or learning rate α_+ (Fig. 2C), suggesting possible generalization. However, only noise/exploration and α_+ predicted age trajectories between tasks, hence revealing deeper similarities, and this was only possible when tasks were sufficiently similar (Table 4), highlighting the limits of generalization. No generalization was possible for α_-, whose age trajectory differed both qualitatively and quantitatively between tasks, showing striking inverse patterns. Like for absolute parameter values, differences in parameter age trajectories were likely caused by differences in task demands. (3) For the parameters whose age trajectories showed signs of generalization, our results replicated patterns in the literature, with noise/exploration decreasing and α_+ increasing from childhood to early adulthood [13].

2.2. Part II: Parameter Interpretability

Tackling the second fundamental assumption we identified above, Part II focuses on parameter interpretability, testing whether parameters captured specific, unique, and meaningful cognitive processes. To this end, we first investigated the relations between different parameters to assess whether individual parameters were uniquely interpretable (i.e., specific and distinct from each other). We then determined how parameters were related to observed behavior, seeking evidence for external interpretability.

2.2.1. Main Axes of Variation

To build the foundation for parameter interpretation, we first aimed to understand which cognitive processes (or other aspects of participant behavior) were captured by each parameter. We opted for a data-driven approach, interpreting parameters based on the major axes of variance that emerged in our large dataset, identified without a priori hypotheses. Concretely, we used PCA to identify the principal components (PCs) of our joint dataset of behavioral features and model parameters. We first gained a thorough understanding of these PCs (summarized below) and then employed them to better understand what was captured by model parameters. Detailed information on our approach is provided in sections 4.6 (PCA methods), Appendix C (behavioral features), and suppl. Fig. D.10 (additional PCA results).

We first analyzed PC1, the axis of largest variation and main source of individual differences in our dataset (25.1% of explained variance; suppl. Fig. D.10A). We found that behaviors that indicated good task participation (e.g.,

higher percentage of correct choices) loaded positively on PC1, whereas behaviors that indicated bad participation loaded negatively (e.g., more missed trials, longer response times; Fig. 3A). This was the case for performance measures in the narrow sense of maximizing choice accuracy (e.g., percentage correct choices, trials to criterion, proportion of win-stay choices), but also in the wider sense of reflecting task engagement (e.g., number of missed trials, response times, response time variability). PC1 therefore captured a range of "good", task-engaged behaviors. PC1 increased significantly with age, consistent with increasing performance (suppl. Fig. B.6B; age effects of subsequent PCs in suppl. Fig. D.10; suppl. Table D.8).

How can this understanding of PC1 help us interpret model parameters? In all three tasks, noise/exploration and Forgetting parameters loaded negatively on PC1 (Fig. 3A), showing that elevated decision stochasticity and decay of previous experiences were associated with poorer performance in all tasks. α_+ showed positive loadings throughout, suggesting that faster integration of positive feedback was associated with better performance in all tasks. Taken together, noise/exploration, Forgetting, and α_+ showed consistency across tasks in terms of their interpretation with respect to behavioral quality. Contrary to this, α_- loaded positively in task C, but negatively in task B, suggesting that performance increased when participants integrated negative feedback faster in task C, but decreased when they did the same in task B. As mentioned before, contradictory patterns of α_- were likely related to task demands, namely the difference in how diagnostic negative feedback was (Fig. 1B). Indeed, this contrasting pattern for α_- was mirrored in "Lose-stay" behavior (repeating choices that produced negative feedback before). Specifically, "lose-stay" loaded positively on PC1 in task B (Fig. 3A), showing it benefited performance in task B, while it loaded negatively on task C, showing that it hurt performance there. This supports the claim that lower α_- was beneficial in task B, while higher α_- was beneficial in task C, in accordance with participants' chosen behavior and developmental differences.

We next analyzed PC2 and PC3. For visualization, we flipped the loadings of all features with negative loadings on PC1 to remove the effects of task engagement (PC1) when interpreting subsequent PCs (for details, see section 4.6). This revealed that PC2 and PC3 encoded task contrasts: PC2 contrasted task B to task C (loadings were positive / negative / near-zero for corresponding features of tasks B / C / A; Fig. 3B). PC3 contrasted task A to both B and C (loadings were positive / negative for correspond-

ing features on task A / tasks B and C; Fig. 3C). (As opposed to most features of our dataset, missed trials and response times did not show these task contrasts, suggesting that these features did not differentiate between tasks). The ordering of PC2 before PC3 shows that participants' behavior differed more between task B compared to C (PC2: 8.9% explained variance) than between B and C compared to A (PC3: 6.2%; suppl. Fig. D.10), as expected based on task similarity (Fig. 1B). PC2 and PC3 hence show that subsequent to task engagement, the main variation in our dataset arose from behavioral differences between tasks.

What can this tell us about parameter interpretation? Intriguingly, noise/exploration parameters, α_+, and α_- reproduced the task contrasts of PC2 and PC3, showing positive or negative loadings based on the task (Fig. 3B, 3C). This means that these parameters differed between tasks and were discriminable (as opposed to, e.g., response times, whose lack of task contrasts suggests consistency). This contradicts the assumption that parameters are specific or interpretable in a task-independent way: If parameters were specific and captured the same cognitive processes across tasks, they would show similar loadings across tasks rather than contrasts.

Taken together, the PCA revealed that the emerging major axes of variation in our large dataset, together capturing 40.2% of explained variance, were task engagement (PC1) and task differences (PC2-PC3). These dimensions can be employed to better understand model parameters: Task engagement (PC1) played a crucial role for all four parameters (Fig. 3A), and this role was consistent across tasks for Noise/exploration, Forgetting, and α_+. This consistency supports the claim that parameters captured specific, task-independent processes in terms of PC1. For α_-, however, PC1 played inverse roles across tasks, showing a lack of task-independent specificity that was likely due to differences in task demands. Furthermore, PC2 and PC3 revealed that Noise/exploration, α_+, and α_- specifically encoded task contrasts, suggesting that the parameters captured different cognitive processes across tasks, lacking a task-independent core of meaning.

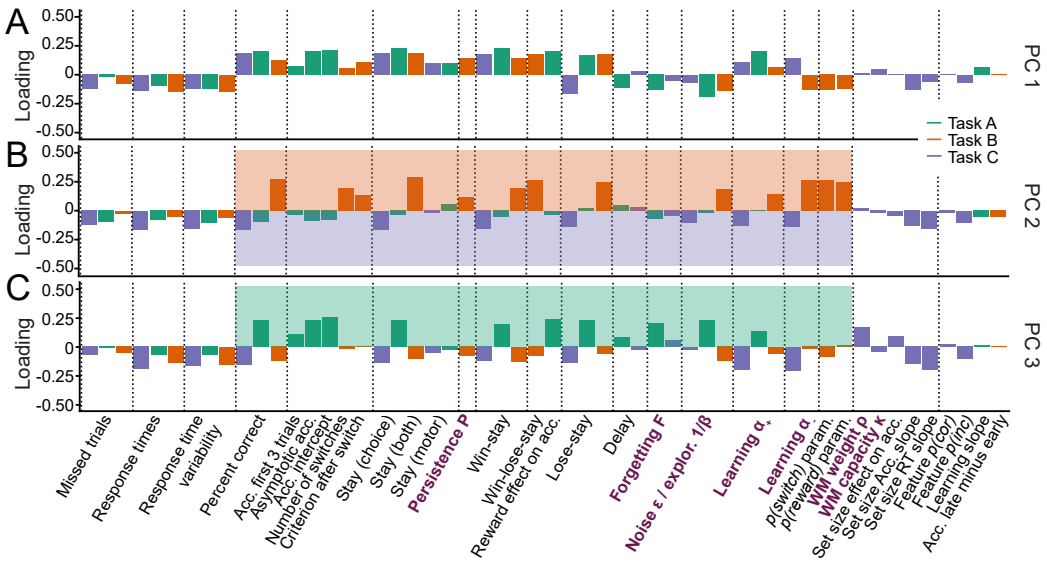

Figure 3: **Identifying the major axes of variation in the dataset.** A PCA was conducted on the entire dataset (39 behavioral features and 15 model parameters). The figure shows the factor loadings of the first three PCs. RL model parameters are highlighted in purple on the x-axis. Behavioral features are explained in detail in Appendix A and Appendix B. (A) PC1 captured broadly-defined task engagement, with negative loadings on features that were negatively associated with performance (e.g., number of missed trials) and positive loadings on features that were positively associated with performance (e.g., percent correct trials). (B-C) PC2 (B) and PC3 (C) captured task contrasts. PC2 loaded positively on features of task B (orange box) and negatively on features of task C (purple box). PC3 loaded positively on features of task A (green box) and negatively on features of tasks B and C. Loadings of features that are negative on PC1 are flipped in PC2 and PC3 to better visualize the task contrasts (section 4.6).

2.2.2. Parameters and Cognitive Processes

Whereas the previous analysis revealed that parameter roles were not entirely consistent across tasks, it did not distinguish between parameter specificity (whether the same parameter captures the same cognitive processes across tasks) and distinctiveness (whether different parameters capture different cognitive processes).

To assess this, we probed how much parameter variance was explained by both corresponding and non-corresponding parameters across tasks, using regression. We assumed that parameters reflected one or more cognitive processes, such that shared variance implies overlapping cognitive pro-

cesses. If parameters are specific (i.e., reflect similar cognitive processes across tasks), then corresponding parameters should be predictive of each other (e.g., predicting task B's $\frac{1}{\beta}$ from task A's parameters, task A's $\frac{1}{\beta}$ should show a significant regression coefficient). If parameters are also distinct, then non-corresponding parameters should furthermore not be predictive (e.g., no other parameters beside task A's $\frac{1}{\beta}$ should predict task B's $\frac{1}{\beta}$). We used repeated, k-fold cross-validated Ridge regression to avoid overfitting, obtaining unbiased out-of-sample estimates of the means and variances of explained variance R^2 and regression coefficients w (for methods, see section 4.7).

Assessing general patterns that arose in this analysis, we found that all significant coefficients connected tasks A and B or tasks A and C but never tasks B and C, mirroring previous results (Fig. 2D; section 2.1.2) with regard to task similarity (Fig. 1B). This suggests that no parameter had a specific core that extended across all three tasks—the largest shared variance encompassed two tasks with high similarity.

Focusing on noise/exploration parameters, coefficients were significant when predicting noise/exploration in task A from noise/exploration in tasks B or C, but the inverse was not true, such that coefficients were not significant when predicting tasks B or C from task A (Fig. 4A; Table 5). The first result implies specificity, showing that noise/exploration parameters captured variance (cognitive processes) in task A that they also captured in tasks B and C. The second result, however, implies a lack of specificity, showing that noise/exploration parameters captured additional cognitive processes in tasks B and C that they did not capture in task A. A further lack of specificity was evident in that even the variance that both B and C captured in A was not the same: Prediction accuracy increased when combining tasks B and C to predict task A, showing that noise/exploration parameters in tasks B and C captured partly non-overlapping aspects of noise/exploration (Fig. 4B, left-most set of bars, compare purple to orange and blue). Assessing distinctiveness, noise/exploration in task A was not only predicted by other noise/exploration parameters, but also by Persistence and α_- in task B, and by α_- and working-memory weight ρ in task C (Fig. 4A; Table 5). This shows that some processes that noise/exploration parameters captured in task A were captured by different parameters in other tasks. In other words, noise/exploration parameters did not capture cognitive processes that were distinct from other parameters.

We next assessed learning rates. Specificity was evident in that learning

rate α_+ in task A showed a significant regression coefficient when predicting learning rates α_+ and α_- in task C, and learning rate α_- in task C showed a significant coefficient when predicting learning rate α_+ in task A (Fig. 4A; Table 5). This suggests a shared core of processing between learning rates α_+ and α_- in tasks A and C. However, a lack of specificity was evident in task B: When predicting α_+ in task B, no parameter of any task showed a significant coefficient (including α_+ in other tasks; Table 5), and it was impossible to predict variance in task B's α_+ even when combining all parameters of the other tasks (Fig. 4B, "Task B" panel). This reveals that α_+ captured fundamentally different cognitive processes in task B compared to the other tasks. The case was similar for parameter α_-, which strikingly was inversely related between tasks A and B (Table 5), and impossible to predict in task B from all other parameters (Fig. 4B). This reveals a fundamental lack of specificity, implying that learning rates in task B did not capture the same core of cognitive processes compared to other tasks.

Regarding distinctiveness, learning rate α_+ in task A was predicted indistinctly by all parameters of task B (with the notable exception of α_+ itself; Fig. 4A; Table 5), suggesting that the cognitive processes that α_+ captured in task A were captured by an interplay of several parameters in task B. Furthermore, task A's α_+ was predicted by task C's working-memory parameters ρ and K (Fig. 4A; Table 5), suggesting that α_+ captured a conglomerate of RL and working-memory processes in task A that was isolated by different parameters in task C [58]. [1] In tasks B and C, α_+ and α_- were partly predicted exclusively by other learning-rate parameters (specific and distinct), partly not predicted at all (lack of specificity), and partly predicted by several parameters (lack of distinctiveness; Fig. 4A). In sum, evidence was mixed with regard to learning rates: Some specificity was evident in the shared variance between tasks A and C, but that specificity was missing in task B. Evidence for distinctiveness was slim, with dominating patterns of variance that were shared between multiple parameters.

[1] In support of this interpretation, no variance in task C's working-memory parameters could be explained by any other parameters (Fig. 4B), suggesting that they captured unique working-memory processes that were not captured by other parameters. Task C's RL parameters, on the other hand, could be explained by parameters in other tasks (Fig. 4B), suggesting they captured overlapping RL processes.

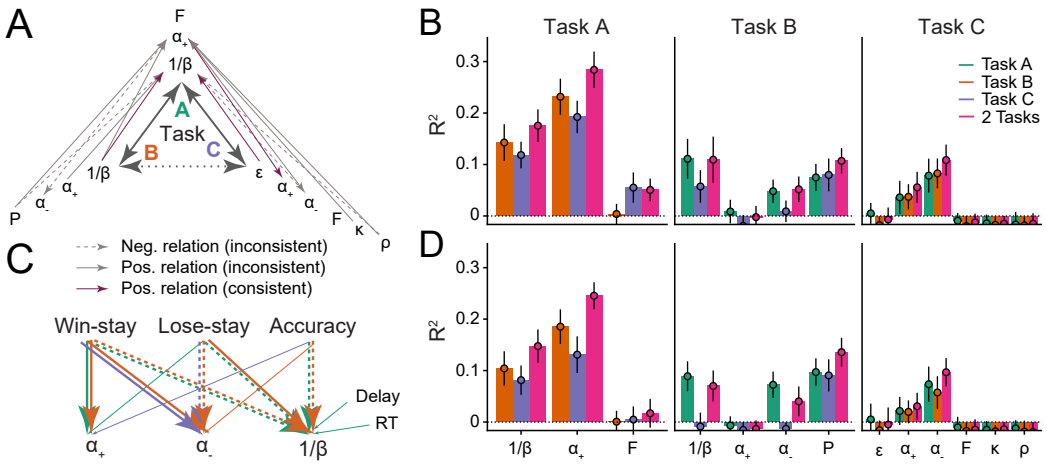

Figure 4: **Assessing parameter interpretability by analyzing shared variance.** (A) Parameter variance that is shared between tasks. Each arrow shows a significant regression coefficient when predicting a parameter in one task (e.g., α_+ in task A) from all parameters of a different tasks (e.g., P, α_-, α_+, and $\frac{1}{\beta}$ in task B). The predicted parameter is shown at the arrow head, predictors at its tail. Full lines indicate positive regression coefficients, and are highlighted in purple when connecting two identical parameters; dotted lines indicate negative coefficients; non-significant coefficients are not shown. Table 5 provides the full statistics of the models summarized in this figure. (B) Amount of variance of each parameter that was captured by parameters of other models. Each bar shows the percentage of explained variance (R^2) when predicting one parameter from all parameters of a different task/model, using Ridge regression. Part (A) of this figure shows the coefficients of these models. The x-axis shows the predicted parameter, and colors differentiate between predicting tasks. Three models were conducted to predict each parameter: One combined the parameters of both other tasks (pink), and two kept them separate (green, orange, blue). Larger amounts of explained variance (e.g., Task A $\frac{1}{\beta}$ and α_-) suggest more shared processes between predicted and predicting parameters; the inability to predict variance (e.g., Task B α_+; Task C working-memory parameters) suggests that distinct processes were captured. Bars show mean R^2, averaged over k data folds (k was chosen for each model based on model fit, using repeated cross-validated Ridge regression; for details, see section 4.7); error bars show standard errors of the mean across folds. (C) Relations between parameters and behavior. The arrows visualize Ridge regression models that predict parameters (bottom row) from behavioral features (top row) within tasks (full statistics in Table 6). Arrows indicate significant regression coefficients, colors denote tasks, and line types denote the sign of the coefficients, like before. All significant within-task coefficients are shown. Task-based consistency (similar relations between behaviors and parameters across tasks) occurs when arrows point from the same behavioral features to the same parameters in different tasks (i.e., parallel arrows). (D) Variance of each parameter that was explained by behavioral features; corresponds to the behavioral Ridge models shown in part (C).

Table 5: Selected coefficients of the repeated, k-fold cross-validated Ridge regression models predicting one parameter from all parameters of a different task. The table includes all significant coefficients and selected non-significant coefficients.

Predicted parameter (Task)	Predicting parameter (Task)	Coefficient	p	sig.
Noise/exploration (A)	Exploration $\frac{1}{\beta}$ (B)	0.14	0.031	*
	α_- (B)	0.14	0.032	*
	Persistence (B)	-0.19	0.0029	**
	Noise ϵ (C)	0.12	0.038	*
	α_- (C)	-0.18	0.045	*
	ρ (C)	-0.19	0.023	*
Noise/exploration (B)	Noise/exploration (A)	0.09	0.27	–
Noise/exploration (C)	Noise/exploration (A)	0.04	0.63	–
α_+ (A)	α_- (C)	0.22	0.011	*
	ρ (C)	0.16	0.050	*
	K (C)	0.15	0.020	*
	Exploration $\frac{1}{\beta}$ (B)	0.19	0.0026	**
	α_- (B)	-0.21	< 0.001	***
	α_+ (B)	0.0042	0.94	–
	Persistence (B)	0.23	< 0.001	***
α_+ (B)	$\frac{1}{\beta}$ (A)	-0.077	0.37	–
	α_+ (A)	0.058	0.48	–
	α_+ (C)	-0.00018	0.99	–
	α_- (C)	-0.000055	1.00	–
	Forgetting (A)	0.015	0.82	–
α_+ (C)	α_+ (A)	0.20	0.013	*
α_- (B)	α_+ (A)	-0.25	0.0018	**
α_- (C)	α_+ (A)	0.24	0.0022	**

2.2.3. Parameters and Behavior

Faced with mounting evidence for parameters' task-dependence and lack of shared core, we lastly aimed to uncover any patterns that were consistent across tasks. The previous sections suggested that parameters captured different cognitive processes across tasks (e.g., different internal characteristics of learning and choice). However, parameters might nevertheless capture the same behavioral features (e.g., similar tendencies to stay after positive feedback). This is possible if the mapping between cognitive processes and behavior is many-to-many, i.e., if different cognitive mechanisms (e.g., reasoning, value learning, episodic memory) can underlie the same behavioral patterns (e.g., lose-stay behavior), and different behavioral patterns (e.g., win-stay,

lose-shift) can be the result of the same cognitive mechanism (e.g, value learning). The nature of these mappings might be modulated by task characteristics (e.g., stable contingencies and/or deterministic feedback might favor reasoning, while volatile contingencies and/or stochastic feedback might favor value learning; Fig. 5B). To investigate this possibility, we assessed the relationships between model parameters and behavioral features across tasks. Using regularized Ridge regression like above, we predicted each model parameter from five behavioral features of each task (15 predictors total, see Appendix A, Appendix C; for regression methods, see section 4.7).

Noise/exploration, indeed, was predicted by the same behavioral features in tasks A and B, such that task A's accuracy, win-stay, and lose-stay behavior predicted task A's $\frac{1}{\beta}$; and task B's accuracy, win-stay, and lose-stay behavior predicted task B's $\frac{1}{\beta}$ (Fig. 4C; Table 6). This shows consistency in terms of which (task-specific) behaviors were related to (task-specific) parameter $\frac{1}{\beta}$. Similarly for learning rates, α_+ was predicted by the same behavior (win-stay) in tasks A and B, and α_- was predicted by the same behaviors (lose-stay, win-stay) in tasks B and C (Fig. 4C; Table 6). The consistency for α_- is especially noteworthy given the abundance of previous discrepancies. These results suggest that parameters successfully generalized between tasks in terms of which behaviors they reflected. Discrepancies in terms of parameter values, age trajectories, and interpretation therefore likely arose because tasks elicited different behaviors, which led to different parameter estimates.

Taken together, noise/exploration parameters, α_+, and α_- captured similar behavioral features across tasks (Fig. 4C), despite differences in cognitive processing (Fig. 4A, 4B), captured variance (Fig. 3B, 3C), age trajectories (Fig. 2C, 2D), and absolute values (Fig. 2A, 2B). Notably, observed discrepancies often made sense in the light of specific task characteristics (Fig. 1B, 3A), both in the case of parameters (Fig. 1B, 3A) and of behavior (suppl. Fig. B.6B), suggesting that both behavioral responses and model parameters were shaped by task characteristics.

Table 6: Statistics of selected coefficients in the repeated, k-fold cross-validated Ridge regression models predicting each model parameter from all behavioral features of all three tasks. The table includes all significant coefficients of within-task predictors, and a selected number of non-significant and between-task coefficients.

Predicted parameter (Task)	Predicting parameter (Task)	coefficient	p	sig.
Noise/exploration (A)	Win-stay (A)	-0.30	< 0.001	***
	Lose-stay (A)	-0.23	< 0.001	***
	Accuracy (A)	-0.19	0.0076	**
	Response times (A)	0.092	0.029	*
	Delay (A)	0.25	< 0.001	***
Noise/exploration (B)	Win-stay (B)	-0.58	< 0.001	***
	Lose-stay (B)	0.091	0.0034	**
	Accuracy (B)	-0.36	< 0.001	***
	Win-stay (A)	-0.12	0.032	*
	Response times (A)	0.059	0.051	–
α_+ (A)	Win-stay (A)	0.74	< 0.001	***
α_+ (B)	Win-stay (B)	0.27	< 0.001	***
α_+ (C)	Accuracy (C)	0.24	0.033	*
α_- (B)	Win-stay (B)	0.29	< 0.001	***
	Lose-stay (B)	-0.71	< 0.001	***
	Accuracy (B)	-0.28	< 0.001	***
α_- (C)	Win-stay (C)	0.16	0.009	**
	Lose-stay (C)	-0.41	< 0.001	***

3. Discussion

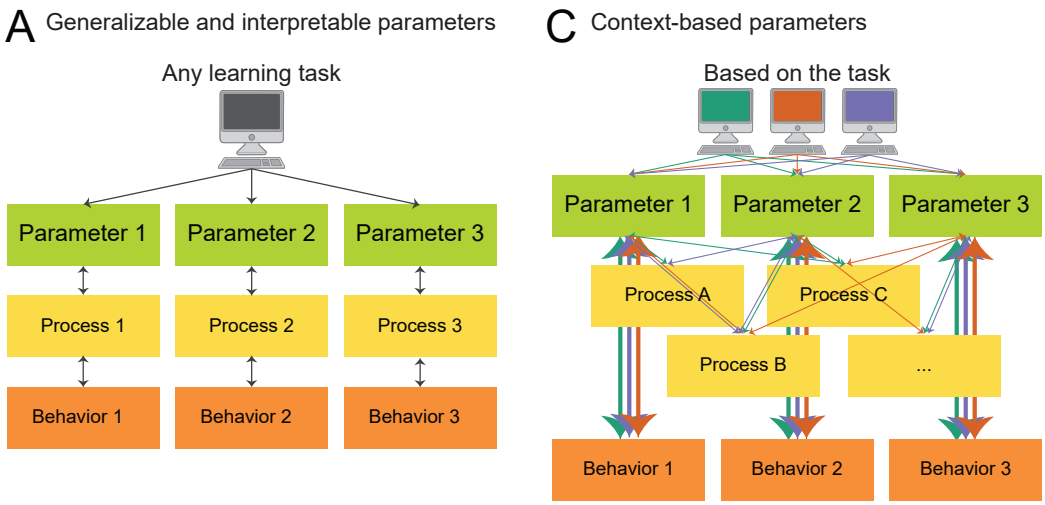

Figure 5: **What do model parameters measure?** (A) View based on generalizability and interpretability. In this view, which is implicitly taken by much current computational modeling research, models are fitted in order to reveal individuals' intrinsic model parameters, which reflect clearly delineated, separable, and meaningful (neuro)cognitive processes, a concept we call *interpretability*. Interpretability is evident in that every model parameter captures a specific cognitive process (bidirectional arrows between each parameter and process), and that cognitive processes are separable from each other (no connections between processes). Task characteristics are treated as irrelevant, a concept we call *generalizability*, such that parameters of any learning task (within reason) are expected to capture similar cognitive processes. (C) Updated view, based on our results, that acknowledges the role of context (e.g., task characteristics, model parameterization, participant sample) in computational modeling. Which cognitive processes are captured by each model parameter is influenced by context (green, orange, blue), as shown by distinct connections between parameters and cognitive processes. Different parameters within the same task can capture overlapping cognitive processes (not interpretable), and the same parameters can capture different processes depending on the task (not generalizable). However, parameters likely capture consistent behavioral features across tasks (thick vertical arrows). Reproduced in part from [28], with permission.

Generalizability has been stated as a major advantage of computational modeling [43], and interpretability (i.e., the inherent "meaningfulness" of parameters) has been a large selling point [14]. The current study tested the generalizability and interpretability of RL models empirically, using a developmental, within-participant approach. Within a set of relatively similar

tasks, we found much weaker levels of generalizability and interpretability than would be expected based on common current research practices, such as comparing parameters between studies with similar—but non-identical—tasks and models [28].

Interestingly, generalizability and interpretability varied between parameters: Exploration/noise parameters showed considerable generalizability in the form of correlated variance and age trajectories, and the decline in exploration/noise we observed between ages 8-17 was consistent with previous studies [13]. Interpretability of exploration/noise parameters, however, was mixed: Despite evidence for specificity in some cases (overlap in parameter variance between some tasks), it was missing in others (lack of overlap between other tasks), and crucially, parameters lacked distinctiveness (substantial overlap in variance with other parameters).

Learning rate from negative feedback showed a substantial lack of generalizability: parameters were less consistent within participants than within tasks, and age trajectories differed both quantitatively and qualitatively. This result is consistent with discrepancies in learning rate parameters across developmental studies [13]. Learning rates from positive and negative feedback combined were specific to a limited degree (overlap in variance between some tasks). However, a lack of specificity (lack of shared core variance) and distinctiveness (fundamental entangling with several other parameters, most notably working-memory parameters) overshadowed this result.

These within-participant findings are consistent with the diverging patterns that are emerging from comparisons of studies published by different labs [13]. Our within-participant design allowed us to go beyond these between-study findings and confirm that the same participants can show different parameters when tested using different tasks. The within-participant consistency of noise/exploration parameters strengthens our confidence that these indeed decrease with age [13, 59, 60]. The inconsistency of learning rate parameters leads to the unexpected, but important conclusion that we cannot measure an individual's "*intrinsic* learning rate" using RL modeling, and that we cannot draw general conclusions about "the development of learning rates" that apply to all RL contexts, using current methods.

These findings help clarify the source of parameter inconsistencies in the previous literature, above and beyond replication problems and technical issues (e.g., model misspecification [13], lack of model comparison and validation [27, 30], lack of model critique [43], inappropriate fitting methods [29, 57], lack of parameter reliability due to suboptimal methods [61]): Our

results show that discrepancies are expected even with a consistent methodological pipeline, and using up-to-date modeling techniques, as an expected consequence of slight task variations. Future research will need to determine the relative contributions of these two sources of inconsistency (technical issues vs expected task-based variation) to date and in the future. Future research is also necessary to study more systematically the external factors that drive such inconsistencies, and are currently undescribed by RL methods.

In sum, our results suggest that relating model parameters to cognitive constructs and real-world behavior might require us to carefully account for task variables and environmental variability in general. This ties into the bigger picture of understanding how neurocognitive processes are shared between tasks [62], and reflects a larger pattern of realization in psychology that we cannot objectively assess an individual's cognitive processing while ignoring subjective context. We have shown that in lab studies, different task contexts recruit different system settings within an individual; similarly, real-life environment, its changes during development, and past environment [63] may also modulate which cognitive processes we recruit. Heightened awareness and systematic study of contextual variables should improve computational modeling of learning and yield valuable new tools to leverage learning across development.

3.1. Limitations

One limitation of our results is that regression analyses might be contaminated by parameter cross-correlations (sections 2.1.2, 2.1.3, 2.1.4), which would reflect modeling limitations (non-orthogonal parameters), and not necessarily shared cognitive processes. For example, parameters α and β are mathematically related in the regular RL modeling framework [26, 29], and we observed significant within-task correlations between these parameters for two of our three tasks (suppl. Fig. D.7). This indicates that caution is required when interpreting correlation results. However, correlations were also present between tasks (suppl. Fig. D.8), suggesting that within-model trade-offs were not the only explanation for shared variance, and that shared cognitive processes likely also played a role. Furthermore, correlations between parameters within models are frequent in the existing literature, and do not prevent researchers from interpreting parameters—in this sense, the existence of similar correlations in our study allows us to address the question of generalizability and interpretability in similar circumstances as in the existing literature.

3.2. Moving Forward

With this research, we do not intend to undermine RL modeling, but to improve its quality. Computational model parameters potentially provide highly valuable insights into (neuro)cognitive processing—we just need to refrain from assuming that the identified processes are—through their mere nature as model parameters—specific, distinct, and "theoretically meaningful" [14]. Parameters with the same names do not necessarily transfer between different tasks of the same family, making them less interchangeable than we often assume [28], whereas the behavioral features captured by parameters might generalize well between tasks.

In the long term, we need to understand why RL parameters differ between tasks. We suggest three potential, not mutually exclusive answers:

1. **Adaptation and Optimality**. Variance in RL parameters may reflect how participants adapt their behavior to task demands, an explanation proposed by [13]. Whereas it is commonly assumed that parameters reflect participants' intrinsic cognitive "settings" (e.g., 10-year-olds have a learning rate of 20%; 16-year-olds of 40%), the optimality-based view suggests that participants instead adaptively tune parameters to task characteristics (e.g., adopting lower learning rates in stable than volatile contexts [55, 64]). Hence, different tasks lead to different parameter estimates because different values are required for optimal behavior; an "optimal" participant—achieving optimal behavior in each task—would therefore naturally show different values across tasks. Similar optimality-based views are held by others [56]. If adaptation to achieve optimality exists, then we would also predict, e..g, that learning rates differ between deterministic and stochastic tasks because both kinds of tasks require different amounts of behavioral change in response to feedback to reach optimal performance. We indeed observed this pattern in the current study. Age differences in parameters can be explained as differences in adaptation flexibility and/or differences in optimal settings due to interaction with different environments (different ages might have different capacities to detect changes or shift internal settings when switching tasks). More research is needed, however, to determine whether parameter optimality and the capacity to optimize behavior can explain all inconsistencies in the literature. For example, our finding that participants showed the most optimal learning rates in the intermediate age range in task B [34], whereas optimality increased

monotonously with age in tasks A and C [33, 35], suggests that other factors besides optimization might play a role as well.

2. **Modulatory processes.** RL Parameters may vary as a function of modulatory processes that are not well captured in current RL models. Modulatory processes have been described in cognition and neurobiology and likely serve to shift functional outputs (e.g., hunger increasing motivation) [65, 66, 67]. Some modulatory processes reflect external contextual information (e.g., uncertainty affects dopamine neuron firing [68, 69, 70]), and RL processes might depend on these same modulatory processes (e.g., RL reward-prediction errors and dopamine [1]). Indeed, environments with different degrees of uncertainty have been shown to elicit different learning rates [55], and EEG markers of neuromodulator release predicted learning rates [71]. It is thus possible that neuromodulation by task uncertainty modulates RL processes, reflected in RL parameters. In our data, feedback stochasticity and task volatility likely contribute to uncertainty-related modulation. However, other factors like task similarity, task volatility [55, 34, 64], feedback stochasticity, memory load [33, 58], feedback valence and conditioning type [24], and choice of model parameters (e.g., forgetting [33, 35], counter-factual learning [34], negative and positive learning rates [37, 72, 73]), have also been shown to affect RL parameters, but are independent of uncertainty. More research is therefore needed to investigate the extent of the contribution of modulatory processes, and its impact on cognition and computation.

3. **RL processes are multifaceted.** RL Parameters capture a multitude of cognitive processes, whose composition likely differs across tasks (Fig. 5B). RL algorithms are framed in the most general way to allow application to a wide range of contexts, including AI, neuroscience, and psychology [26, 28, 74, 75]. As behavioral models, their use has spanned a variety of behaviors, meaning that the same parameters capture cognitive processes that vary considerably in type and complexity: The same RL parameters have been said to capture the slow acquisition of implicit preferences [1], long-term memory for preferences [76], quick recognition of contingency switches [34, 77], selection of abstract high-level strategies [7, 9, 78], meta-learning [6], habitual and goal-directed decision making [5], working memory or episodic memory-guided choice [58, 79, 80], and many others. This list of cognitive processes outnumbers the list of typical RL model parameters, suggesting that RL pa-

rameters necessarily capture different (combinations of) cognitive processes depending on the paradigm. Indeed, adaptive learning does not seem to be a unitary phenomenon, but seems to be composed of several distinct neuro-cognitive factors [56]. Similar arguments have been made for behavioral analyses [81].

3.3. Conclusion

Our research has important implications for computational modeling in general, and specifically for fields that focus on individual differences, including developmental and clinical computational research: We show that contextual factors critically impact computational modeling results. Larger, targeted studies will be necessary to identify the most important contextual factors and their precise roles, and to quantify their effects. Other areas of modeling besides RL should be subjected to a similar investigation. It is possible, for example, that generalizability and interpretability are also overstated in models of sequential sampling [82, 83], Bayesian inference [34, 84, 85], model-based versus model-free RL [61, 86, 87], or others.

A lack of generalizability and interpretability means that model parameters do not measure task-independent, person-specific characteristics, as we often assume. Instead, parameters might be much more closely tied to the specific, contextual factors of experimental paradigms and also the historical context brought by the subject. Acknowledging this should help the field of computational modeling to accurately interpret individual computational models (in direct relation to the experimental task), to combine insights of different studies (by taking into account differences in parameter optimality, modulatory factors, and processes captured by each parameter), and to achieve actual generalizability and interpretability of findings in the future. In the long term, these findings should not undercut the field, but help it to become more robust.

4. Methods

4.1. Study Design

Our sample of 291 participants was balanced between females and males, and all ages (8-30 years) were represented equally (Fig. 1A, left). Participants completed four computerized tasks, questionnaires, and a saliva sample during the 1-2 hour lab visit (see section 4.3). To reduce noise, we excluded participants based on task-specific performance criteria (see section 4.2). Due

to worse performance, more younger than older participants were excluded, which is a caveat for the interpretation of age effects (note however that these exclusions cannot account for the observed age effects but act against them; Fig. 1A). Our tasks—A ("Butterfly task" [35, 48]), B ("Stochastic Reversal" [77, 34]), and C ("Reinforcement learning-Working memory" [33, 58])—were all classic reinforcement learning tasks: on each trial, participants chose between several actions in an effort to earn rewards, which were presented as binary feedback (win/point or lose/no point) after each choice.

The tasks varied on several common dimensions (Fig. 1B), which have been related to discrepancies in behavioral and neurocognitive results in the literature [24, 38, 39]. For example, in one task (task C), positive feedback was deterministic, such that every correct action led to a positive outcome, whereas in the two other tasks (tasks A and B), positive feedback was stochastic, such that some correct actions led to positive and others to negative outcomes. A different set of two tasks (B and C) provided diagnostic positive feedback, such that every positive outcome indicated a correct action, whereas in the third (A), positive feedback was non-diagnostic, such that positive outcomes could indicate both correct and incorrect actions. Two tasks (A and C) presented several different stimuli/states for which correct actions had to be learned, whereas the third (B) only presented a single one. Overall, task A shared more important similarities with both tasks B and C than either of these shared with each other, allowing us to explore whether task similarity played a role in parameter generalizability and interpretability. A comprehensive list of task differences is shown in Fig. 1B, and each task is described in more detail in section 4.4. Section Appendix B explains the most prominent findings of each task individually, and shows several behavioral measures over age.

4.2. Participant Sample

4.2.1. Sample Overview

All procedures were approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley. We tested 312 participants: 191 children and adolescents (ages 8-17) and 55 adults (ages 25-30) were recruited from the community and completed a battery of computerized tasks, questionnaires, and saliva samples; 66 university undergraduate students (aged 18-50) completed the four tasks as well, but not the questionnaires or saliva sample. Community participants of all ages were prescreened for the absence of present or past psychological and neurological disorders;

the undergraduate sample indicated the absence of these. Compensation for community participants consisted in $25 for the 1-2 hour in-lab portion of the experiment and $25 for completing optional take-home saliva samples; undergraduate students received course credit for participation in the 1-hour study.

4.2.2. Participant Exclusion

Two participants from the undergraduate sample were excluded because they were older than 30, and 7 were excluded because they failed to indicate their age. This led to a sample of 191 community participants under 18, 57 undergraduate participants between the ages of 18-28, and 55 community participants between the ages of 25-30. Of the 191 participants under 18, 184 completed task B, and 187 completed tasks A and C. Reasons for not completing a task included getting tired, running out of time, and technical issues. All 57 undergraduate participants completed tasks B and C and 55 completed task A. All 55 community adults completed tasks B and A, and 45 completed task C. Appropriate exclusion criteria were implemented separately for each task to exclude participants who failed to pay attention and who performed critically worse than the remaining sample (for task A, see [35]; task B [34]; task C [33]). Based on these criteria, 5 participants under the age of 18 were excluded from task B, 10 from task A, and none from task C. One community adult participant was excluded from task A, but no adult undergraduates or community participants were excluded from tasks B or C.

Because this study related the results across all three tasks, we only included participants who were not excluded in any task, leading to a final sample of 143 participants under the age of 18 (male: 77; female: 66), 51 undergraduate participants (male: 17; female: 34), and 53 adults from the community (male: 25; female: 28), for a total of 247 participants (male: 119; female: 128). We entirely excluded the fourth task of our study from the current analysis, which was modeled after a rodent task and used in humans for the first time [88], because the applied performance criterion led to the exclusion of the majority of our developmental sample. We split participants into quantiles based on age, which were calculated separately within each sex (for details, see [34]).

4.3. Testing Procedure

After entering the testing room, participants under 18 years and their guardians provided informed assent and permission; participants over 18 provided informed consent. Guardians and participants over 18 filled out a demographic form. Participants were led into a quiet testing room in view of their guardians, where they used a video game controller to complete four computerized tasks. The first task was called "4-choice" and assessed reversal learning in an environment with 4 different choice options, with a duration of approximately 5 minutes (designed after [88]). This task was excluded from the current analysis (see section 4.2.2). The second task was C ("Reinforcement learning-Working memory") and took about 25 minutes to complete [58, 33]. After the second task, participants between the ages of 8-17 provided a saliva sample (for details, see [33]) and took a snack break (5-10 minutes). After that, participants completed task A ("Butterfly task"), which took about 15 minutes [48, 35], and task B ("Stochastic Reversal"), which took about 10 minutes to complete [34]. At the conclusion of the tasks, participants between 11 and 18 completed the Pubertal Development Scale (PDS [89]) and were measured in height and weight. Participants were then compensated with $25 Amazon gift cards. The PDS questionnaire and saliva samples were administered to investigate the role of pubertal maturation on learning and decision making. Pubertal analyses are not the focus of the current study and will be or have reported elsewhere [33, 34, 35]. For methodological details, refer to [33]. The entire lab visit took 60-120 minutes, depending on the participant.

4.4. Task Design

4.4.1. Task A ("Butterfly task")

The goal of task A was to collect as many points as possible, by guessing correctly which of two flowers was associated with each of four butterflies. Correct guesses were rewarded with 70% probability, and incorrect guesses with 30%. The task contained 120 trials (30 for each butterfly) that were split into 4 equal-sized blocks, and took between 10-20 minutes to complete. More detailed information about methods and results can be found in [35].

4.4.2. Task B ("Stochastic Reversal")

The goal of task B was to collect golden coins, which were hidden in two green boxes. The task could be in one of two states: "Left box is correct" or "Right box is correct". In the former, selecting the left box led to reward

in 75% of trials, while selecting the right box never led to a reward (0%). Several times throughout the task, task contingencies changed unpredictably and without notice (after participants had reached a performance criterion indicating they had learned the current state), and the task switched states. Participants completed 120 trials of this task (2-9 reversals), which took approximately 5-15 minutes. For more information and additional task details, refer to [34].

4.4.3. Task C ("Reinforcement Learning-Working Memory")

The goal of task C was to collect as many points as possible by pressing the correct key for each stimulus. Pressing the correct key deterministically led to reward, and the correct key for a stimulus never changed. Stimuli appeared in blocks that varied in the number of different stimuli, with set sizes ranging from 2-5. In each block, each stimulus was presented 12-14 times, for a total of 13 * set size trials per block. Three blocks were presented for set sizes 2-3, and 2 blocks were presented for set sizes 4-5, for a total of 10 blocks. The task took between 15-25 minutes to complete. For more details, as well as a full analysis of this dataset, refer to [33].

4.5. Computational Models

For all tasks, we used RL theory to model how participants adapted their behavior in order to maximize reward. RL models assume that agents learn a policy $\pi(a|s)$ that determines (probabilistically) which action a to take in each state s of the world [26]. Here and in most cognitive RL models, this policy is based on action values $Q(a|s)$, i.e., the values of each action a in each state s. Agents learn action values by observing the reward outcomes, r_t, of their actions at each time step t. Learning consists in updating existing action values $Q_t(a|s)$ using the "reward prediction error", the difference between the expected reward $Q_t(a|s)$ and the actual reward r_t:

$$Q_{t+1}(a|s) = Q_t(a|s) + \alpha(r_t - Q_t(a|s))$$

How much a learner weighs past action value estimates compared to new outcomes is determined by parameter α, the learning rate. Small learning rates favor past experience and lead to stable learning over long time horizons, while large learning rates favor new outcomes and allow for faster and more flexible changes, focusing on shorter time horizons. With enough time and

in a stable environment, the RL updating scheme guarantees that value estimates will reflect the environment's true reward probabilities, and thereby allow for optimal long-term choices [26].

In order to choose actions, most cognitive RL models use a (noisy) "softmax" function to translate action values $Q(a|s)$ into policies $p(a|s)$:

$$p(a_i|s) = \frac{exp(\beta\ Q(a_i|s))}{\sum_{a_j \in A} exp(\beta\ Q(a_j|s))}$$

A refers to the set of all available actions (tasks A and B have 2 actions, task C has 3), and a_i and a_j to individual actions within the set. How deterministically versus noisily this translation is executed is determined by exploration parameters β, also called inverse decision temperature, and/or ϵ, the decision noise (see below). Small decision temperatures $\frac{1}{\beta}$ favor the selection of the highest-valued actions, enabling exploitation, whereas large decision temperatures select actions of low and high values more evenly, enabling exploration. Parameter ϵ adds undirected noise to action selection, selecting random action with a small probability ϵ on each trial.

Besides α, β, and noise, cognitive RL models often include additional parameters to better fit empirical behavior in humans or animals. Common choices include Forgetting—a consistent decay of action values back to baseline—, and Persistence—the tendency to repeat the same action independent of outcomes, a parameter also known as sticky choice or perseverance [73]. In addition, cognitive models often differentiate learning from positive versus negative rewards, splitting learning rate α into two separate parameters α_+ and α_-, which are applied to only positive and only negative outcomes, respectively [37, 44, 90, 91, 92, 93, 50, 94, 95]. The next paragraphs introduce these parameters in detail.

In task A, the best fitting model included a forgetting mechanism, which was implemented as a decay in Q-values applied to all action values of the three stimuli (butterflies) that were not shown on the current trial:

$$Q_{t+1}(a|s) = (1 - f) * Q_{t+1}(a|s) + f * 0.5.$$

The free parameter $0 < f < 1$ reflects individuals' tendencies to forget.

In task B, free parameter P captured choice persistence, which biased choices on the subsequent trial toward staying ($P > 0$) or switching ($P < 0$). P modifies action values $Q(a|s)$ into $Q'(a|s)$, as follows:

$$Q'_t(a|s) = Q_t(a|s) + P \iff a_t = a_{t-1}$$

$$Q'_t(a|s) = Q_t(a|s) \iff a_t \neq a_{t-1}$$

In addition, the model of task B included counter-factual learning parameters α_{C+} and α_{C-}, which added counter-factual updates based on the inverse outcome and affected the non-chosen action. For example, after receiving a positive outcome ($r = 1$) for choosing left (a), counter-factual updating would lead to an "imaginary" negative outcome ($\bar{r} = 0$) for choosing right ($\bar{a}$).

$$Q_{t+1}(\bar{a}|s) = Q_t(\bar{a}|s) + \alpha_{C+}(\bar{r} - Q_t(\bar{a}|s)) \iff r = 1$$

$$Q_{t+1}(\bar{a}|s) = Q_t(\bar{a}|s) + \alpha_{C-}(\bar{r} - Q_t(\bar{a}|s)) \iff r = 0$$

$\bar{a}$ indicates the non-chosen action, and $\bar{r}$ indicates the inverse of the received outcome, $\bar{r} = 1 - r$. The best model fits were achieved with $\alpha_{C+} = \alpha_+$ and $\alpha_{C-} = \alpha_-$, so counter-factual learning rates are not reported in this paper.

In tasks A and B, positive and negative learning rates are differentiated in the following way:

$$Q_{t+1}(a|s) = Q_t(a|s) + \alpha_+(r_t - Q_t(a|s)) \iff r_t = 1$$

$$Q_{t+1}(a|s) = Q_t(a|s) + \alpha_-(r_t - Q_t(a|s)) \iff r_t = 0$$

In the best model for task A, only α_+ was a free parameter, while α_- was fixed to 0. In task C, α_- was a function of α_+, such that $\alpha_- = b * \alpha_+$, where b is the neglect bias parameter that determines how much negative feedback is neglected compared to positive feedback. Throughout the paper, we report $\alpha_- = b * \alpha_+$ for task C.

In addition to an RL module, the model of task C included a working-memory module with perfect recall of recent outcomes, but subject to forgetting and capacity limitations. Perfect recall was modeled as an RL process with learning rate $\alpha_{WM+} = 1$ that operated on working-memory weights $W(a|s)$ rather than action values. On trials with positive outcomes ($r = 1$), the model reduces to:

$$W_{t+1}(a|s) = r_t$$

On trials with negative outcomes ($r = 0$), multiplying $\alpha_{WM+} = 1$ with the neglect bias b leads to potentially less-than perfect memory:

$$W_{t+1}(a|s) = W_t(a|s) + b * (r_t - W_t(a|s))$$

Working-memory weights $W(a|s)$ were transformed into action policies $p_{WM}(a|s)$ in a similar way as RL weights $Q(a|s)$ were transformed into action probabilities $p_{RL}(a|s)$, using a softmax transform combined with undirected noise:

$$p(a_i|s) = (1-\epsilon) * \frac{exp(\beta\ Q(a_i|s))}{\sum_{a_j \in a} exp(\beta\ Q(a_j|s))} + \epsilon * \frac{1}{|a|}$$

$|a| = 3$ is the number of available actions and $\frac{1}{|a|}$ is the uniform policy over these actions; ϵ is the undirected noise parameter.

Forgetting was implemented as a decay in working-memory weights $W(a|s)$ (but not RL Q-values):

$$W_{t+1}(a|s)_{t+1} = (1-f) * W_t(a|s)_t + f * \frac{1}{3}$$

Capacity limitations of working memory were modeled as an adjustment in the weight w of $p_{WM}(a|s)$ compared to $p_{RL}(a|s)$ in the final calculation of action probabilities $p(a|s)$:

$$w = \rho * (min(1, \frac{K}{ns}))$$

$$p(a|s) = w * p_{WM}(a|s) + (1-w) * p_{RL}(a|s)$$

The free parameter ρ is the individual weight of working memory compared to RL, ns indicates a block's stimulus set size, and K captures individual differences in working-memory capacity.

We fitted a separate RL model to each task, using state-of-the-art methods for model construction, fitting, and validation [30, 27]. Models for tasks A and B were fitted using hierarchical Bayesian methods with Markov-Chain Monte-Carlos sampling, which is an improved method compared to maximum likelihood that leads to better parameter recovery, amongst other advantages [96, 97, 98]. The model for task C was fitted using classic non-hierarchical maximum-likelihood because model parameter K is discrete, which renders hierarchical sampling less tractable. In all cases, we verified that the model parameters were recoverable by the selected model-fitting procedure, and that the models were identifiable. Details of model-fitting procedures are provided in the original publications [33, 34, 35].

For additional details on any of these models, as well as detailed model comparison and validation, the reader is referred to the original publications (task A: [35]; task B: [34]; task C: [33]).

4.6. Principal Component Analysis (PCA)

The PCA in section 2.2.1 included 15 model parameters (α_+ and noise/exploration in each task; Forgetting and α_- in two tasks; Persistence in task B; four working-memory parameters in task C; see section 4.5) and 39 model-free features, including simple behavioral features (e.g., overall performance, reaction times, tendency to switch), results of behavioral regression models (e.g., effect of stimulus delay on accuracy), and the model parameters of an alternative Bayesian inference model in task B. All behavioral features, including their development over age, are described in detail in Appendix C and suppl. Fig. B.6B. For simplicity, section 2.2.1 focused on the first three PCs only; the weights, explained variance, and age trajectories of remaining PCs are shown in suppl. Fig. D.10.

PCA is a statistical tool that decomposes the variance of a dataset into so-called "principal components" (PCs). PCs are linear combinations of a dataset's original features (e.g., response times, accuracy, learning rates), and explain the same variance in the dataset as the original features. The advantage of PCs is that they are orthogonal to each other and therefore capture independent aspects of the data. In addition, subsequent PCs explain subsequently less variance, such that selecting just the top PCs of a dataset retains the bulk of the variance and the ability to reconstruct the dataset up to random noise. When using this approach, it is important to understand which concept each PC captures. So-called factor loadings, the original features' weights on each PC, can provide this information.

PCA performs a *change of basis*: Instead of describing the dataset using the original features (in our case, 54 behaviors and model parameters), it creates new features, PCs, that are linear combinations of the original features and capture the same variance, but are orthogonal to each other. PCs are created by eigendecomposition of the covariance matrix of the dataset: the eigenvector with the largest eigenvalue shows the direction in the dataset in which most variance occurs, and represents the first PC. Eigenvectors with subsequently smaller eigenvalues form subsequent PCs. PCA is related to Factor analysis, and often used for dimensionality reduction. In this case, only a small number of PCs is retained whereas the majority is discarded, in an effort to retain most variance with a reduced number of features.

We highlight the most central behavioral features here; more detail is provided in Appendix A and Appendix C. Response to feedback was assessed using features "Win-stay" (percentage of trials in which a rewarded choice was repeated), and "Lose-stay" (percentage of trials in which a non-rewarded

choice was repeated). For task B, we additionally included "Win-lose-stay" tendencies, which is the proportion of trials in which participants stay after a winning trial that is followed by a losing trial. This is an important measure for this task because the optimal strategy required staying after single losses.

We also included behavioral persistence measures in all tasks. In tasks A and C, these included a measure of action repetition (percentage of trials in which the previous key was pressed again, irrespective of the stimulus and feedback) and choice repetition (percentage of trials in which the action was repeated that was previously selected for the same stimulus, irrespective of feedback). In task B, both measures were identical because every trial presents the same stimulus.

We further included task-specific measures of performance. In task A, these were: the average accuracy for the first three presentations of each stimulus, reflecting early learning speed; and the asymptote, intercept, and slope of the learning progress in a regression model predicting performance (for details about these measures, see [35]. In task B, task-specific measures of performance included the number of reversals (because reversals were performance-based); and the average number of trials to reach criterion after a switch. In tasks A and C, we also included a model-independent measure of forgetting. In task A, this was the effect of delay on performance in the regression model mentioned above. In task C, this was the effect of delay in a similar regression model, which also included set size, the number of previous correct choices, and the number of previous incorrect choices, whose effects were also included. Lastly for task C, we included the slope of accuracy and response times over set sizes, as measures of the effect of set size on performance. For task B, we also included the difference between early (first third of trials) and late (last third) performance as a measure of learning. To avoid biases in the PCA toward any specific task, we included equal numbers of behavioral features for each task.

To facilitate the interpretation of PC2 and PC3, we normalized the loadings (PCA weights) of each feature (behavioral and model parameter) with respect to PC1, flipping the loadings of all features in PC2 and PC3 that loaded negatively on PC1. This step ensured that the directions of factor loadings on PC2 and PC3 were interpretable in the same way for all features, irrespective of their role for task performance, and revealed the encoding of task contrasts.

4.7. Ridge Regression

In sections 2.2.2 and 2.2.3, we use regularized, cross-validated Ridge regression to determine whether parameters captured overlapping variance, which would point to an overlap in cognitive processes. We used Ridge regression to avoid problems that would be caused by overfitting when using regular regression models. Ridge regression regularizes regression weight parameters w based on their L2-norm. Regular regression identifies a vector of regression weights w that minimize the linear least squares $||y - wX||_2^2$. Here, $||a||_2^2 = \sqrt{\sum_{a_i \in x} a_i^2}$ is the L2-norm of a vector a, vector y represents the outcome variable (in our case, a vector of parameters, one fitted to each participant), matrix X represents the predictor variables (in our case, either several behavioral features for each participant [2.2.2], or several parameters fitted to each participant 2.2.3]), and vector w represents the weights assigned to each feature in X (in our case, the weight assigned to each predicting behavioral pattern or each predicting parameter).

When datasets are small compared to the number of predictors in a regression model, *exploding* regression weights w can lead to overfitting. Ridge regression avoids this issue by not only minimizing the linear least squares like regular regression, but also the L2 norm of weights w, i.e., by minimizing $||y - wX||_2^2 + \alpha * ||w||_2^2$. Parameter α is a hyper-parameter of Ridge regression, which needs to be chosen by the experimenter. To avoid bias in the selection of α, we employed repeated cross-validated grid search. At each iteration of this procedure, we split the dataset into a predetermined number $s \in [2, 3, ..., 8]$ of equal-sized folds, and then fitted a Ridge regression to each fold, using values of $\alpha \in [0, 10, 30, 50, 100, 300, ..., 10{,}000, 100{,}000, 1{,}000{,}000]$. For each s, we determined the best value of α based on cross-validation between folds, using the amount of explained variance, R^2, as the selection criterion. To avoid biases based on the random assignment of participants into folds, we repeated this procedure $n = 100$ times for each value of α. To avoid biases due to the number of folds, the entire process was repeated for each s, and the final value of s was selected based on R^2. We used the python package "scikit learn" [99] to implement the procedure.

We conducted three models per parameter to determine the relations between parameters: predicting each parameter from all the parameters of each of the other two tasks (2 models); and predicting each parameter from all parameters of both other tasks combined (1 model; Fig. 4A). We conducted the same three models per parameter to determine the relations between pa-

rameters and behaviors, predicting each parameter from behavioral features of the other tasks (Fig. 4A). In addition, we conducted a fourth model for behaviors, predicting each parameter from the behaviors of all three tasks combined, to assess the contributions of all behaviors to each parameter (Fig. 4C). Meta-parameters s and α were allowed to differ (and differed) between models. The final values of R^2 (Fig. 4B and 4D) and the final regression weights w (Fig. 4A and 4C; Table 6) were determined by refitting the winning model.

5. Acknowledgments

We thank Ian Ballard, Mayank Agrawal, Gautam Agarwal, and Bas van Opheusden for helpful comments on this manuscript, and Catherine Hartley and other members of her lab for fruitful discussion. Numerous people contributed to this research: Amy Zou, Lance Kriegsfeld, Celia Ford, Jennifer Pfeifer, Megan Johnson, Vy Pham, Rachel Arsenault, Josephine Christon, Shoshana Edelman, Lucy Eletel, Neta Gotlieb, Haley Keglovits, Julie Liu, Justin Morillo, Nithya Rajakumar, Nick Spence, Tanya Smith, Benjamin Tang, Talia Welte, and Lucy Whitmore. We are also grateful to our participants and their families. The work was funded by National Science Foundation SL-CN grant 1640885 to RD, AGEC, and LW.

References

[1] W. Schultz, P. Dayan, P. R. Montague, A Neural Substrate of Prediction and Reward, Science 275 (5306) (1997) 1593–1599. doi:10.1126/science.275.5306.1593.

[2] J. P. O'Doherty, P. Dayan, J. Schultz, R. Deichmann, K. Friston, R. J. Dolan, Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning, Science 304 (5669) (2004) 452–454, publisher: American Association for the Advancement of Science Section: Report. doi:10.1126/science.1094285.
URL `https://science.sciencemag.org/content/304/5669/452`

[3] J. Gläscher, A. N. Hampton, J. P. O'Doherty, Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making, Cerebral Cortex (New York, N.Y.: 1991) 19 (2) (2009) 483–495. doi:10.1093/cercor/bhn098.

[4] J. Ribas Fernandes, A. Solway, C. Diuk, J. T. McGuire, A. G. Barto, Y. Niv, M. Botvinick, A Neural Signature of Hierarchical Reinforcement Learning, Neuron 71 (2) (2011) 370–379. doi:10.1016/j.neuron.2011.05.042.

[5] N. Daw, S. Gershman, B. Seymour, P. Dayan, R. Dolan, Model-Based Influences on Humans' Choices and Striatal Prediction Errors, Neuron 69 (6) (2011) 1204–1215. doi:10.1016/j.neuron.2011.02.027.

[6] J. X. Wang, Z. Kurth-Nelson, D. Kumaran, D. Tirumala, H. Soyer, J. Z. Leibo, D. Hassabis, M. Botvinick, Prefrontal cortex as a meta-reinforcement learning system, Nature Neuroscience 21 (6) (2018) 860–868. doi:10.1038/s41593-018-0147-8.

[7] M. K. Eckstein, A. G. E. Collins, Computational evidence for hierarchically structured reinforcement learning in humans, Proceedings of the National Academy of Sciences 117 (47) (2020) 29381–29389. doi:10.1073/pnas.1912330117.
URL `https://www.pnas.org/content/117/47/29381`

[8] M. Botvinick, Hierarchical reinforcement learning and decision making, Current Opinion in Neurobiology 22 (6) (2012) 956–962. doi:10.1016/j.conb.2012.05.008.
URL `http://linkinghub.elsevier.com/retrieve/pii/S0959438812000876`

[9] A. G. E. Collins, E. Koechlin, Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making, PLOS Biology 10 (3) (2012) e1001293. doi:10.1371/journal.pbio.1001293.

[10] D. M. Werchan, A. G. E. Collins, M. J. Frank, D. Amso, Role of Prefrontal Cortex in Learning and Generalizing Hierarchical Rules in 8-Month-Old Infants, The Journal of Neuroscience 36 (40) (2016) 10314–10322. doi:10.1523/JNEUROSCI.1351-16.2016.
URL `http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1351-16.2016`

[11] W. van den Bos, R. Bruckner, M. R. Nassar, R. Mata, B. Eppinger, Computational neuroscience across the lifespan: Promises and pitfalls, Developmental Cognitive Neuroscience (Oct. 2017). doi:10.1016/j.dcn.2017.09.008.
URL `http://linkinghub.elsevier.com/retrieve/pii/S1878929317301068`

[12] F. Bolenz, A. M. F. Reiter, B. Eppinger, Developmental Changes in Learning: Computational Mechanisms and Social Influences, Frontiers in Psychology 8, publisher: Frontiers (2017). doi:10.3389/fpsyg.2017.02048.
URL https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02048/full

[13] K. Nussenbaum, C. A. Hartley, Reinforcement learning across development: What insights can we draw from a decade of research?, Developmental Cognitive Neuroscience 40 (2019) 100733. doi:10.1016/j.dcn.2019.100733.
URL http://www.sciencedirect.com/science/article/pii/S1878929319303202

[14] Q. J. M. Huys, T. V. Maia, M. J. Frank, Computational psychiatry as a bridge from neuroscience to clinical applications, Nature neuroscience 19 (3) (2016) 404–413. doi:10.1038/nn.4238.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5443409/

[15] R. A. Adams, Q. J. M. Huys, J. P. Roiser, Computational Psychiatry: towards a mathematically informed understanding of mental illness, Journal of Neurology, Neurosurgery & Psychiatry 87 (1) (2016) 53–63, publisher: BMJ Publishing Group Ltd Section: Neuropsychiatry. doi:10.1136/jnnp-2015-310737.
URL https://jnnp.bmj.com/content/87/1/53

[16] T. U. Hauser, G.-J. Will, M. Dubois, R. J. Dolan, Annual Research Review: Developmental computational psychiatry, Journal of Child Psychology and Psychiatry 60 (4) (2019) 412–426. doi:https://doi.org/10.1111/jcpp.12964.

[17] W.-Y. Ahn, J. R. Busemeyer, Challenges and promises for translating computational tools into clinical practice, Current Opinion in Behavioral Sciences 11 (2016) 1–7. doi:10.1016/j.cobeha.2016.02.001.
URL https://www.sciencedirect.com/science/article/pii/S2352154616300237

[18] L. Deserno, R. Boehme, A. Heinz, F. Schlagenhauf, Reinforcement Learning and Dopamine in Schizophrenia: Dimensions of Symptoms or Specific Features of a Disease Group?, Frontiers in Psychiatry 4, publisher: Frontiers (2013). doi:10.3389/fpsyt.2013.00172.
URL https://www.frontiersin.org/articles/10.3389/fpsyt.2013.00172/full

[19] M. J. Frank, E. D. Claus, Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal., Psychological Review 113 (2) (2006) 300–326. doi:10.1037/0033-295X.113.2.300.
URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.113.2.300

[20] Y. Niv, Reinforcement learning in the brain, Journal of Mathematical Psychology 53 (3) (2009) 139–154.

[21] D. Lee, H. Seo, M. W. Jung, Neural Basis of Reinforcement Learning and Decision Making, Annual review of neuroscience 35 (2012) 287–308. doi:10.1146/annurev-neuro-062111-150512.

[22] J. P. O'Doherty, S. W. Lee, D. McNamee, The structure of reinforcement-learning mechanisms in the human brain, Current Opinion in Behavioral Sciences 1 (2015) 94–100. doi:10.1016/j.cobeha.2014.10.004.

[23] P. W. Glimcher, Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis, Proceedings of the National Academy of Sciences 108 (3) (2011) 15647–15654.

[24] J. Garrison, B. Erdeniz, J. Done, Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies, Neuroscience & Biobehavioral Reviews 37 (7) (2013) 1297–1310. doi:10.1016/j.neubiorev.2013.03.023.
URL http://www.sciencedirect.com/science/article/pii/S0149763413000833

[25] P. Dayan, Y. Niv, Reinforcement learning: The Good, The Bad and The Ugly, Current Opinion in Neurobiology 18 (2) (2008) 185–196. doi:10.1016/j.conb.2008.08.003.
URL https://linkinghub.elsevier.com/retrieve/pii/S0959438808000767

[26] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, 2nd Edition, MIT Press, Cambridge, MA; London, England, 2017.

[27] S. Palminteri, V. Wyart, E. Koechlin, The Importance of Falsification in Computational Cognitive Modeling, Trends in Cognitive Sciences 21 (6) (2017) 425–433. doi:10.1016/j.tics.2017.03.011.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364661317300542

[28] M. K. Eckstein, L. Wilbrecht, A. G. Collins, What do reinforcement learning models measure? Interpreting model parameters in cognition and neuroscience, Current Opinion in Behavioral Sciences 41 (2021) 128–137. doi:10.1016/j.cobeha.2021.06.004.
URL https://www.sciencedirect.com/science/article/pii/S2352154621001236

[29] N. D. Daw, Trial-by-trial data analysis using computational models, Decision Making, Affect, and Learning: Attention and Performance XXIII (2011). doi:10.1093/acprof:oso/9780199600434.003.0001.

[30] R. C. Wilson, A. G. Collins, Ten simple rules for the computational modeling of behavioral data, eLife 8 (2019) e49547, publisher: eLife Sciences Publications, Ltd. doi:10.7554/eLife.49547.
URL https://doi.org/10.7554/eLife.49547

[31] O. Guest, A. E. Martin, How Computational Modeling Can Force Theory Building in Psychological Science, Perspectives on Psychological Science (2021) 1745691620970585Publisher: SAGE Publications Inc. doi:10.1177/1745691620970585.
URL https://doi.org/10.1177/1745691620970585

[32] G. Blohm, K. P. Kording, P. R. Schrater, A How-to-Model Guide for Neuroscience, eNeuro 7 (1), publisher: Society for Neuroscience Section: Research Article: Methods/New Tools (Jan. 2020). doi:10.1523/ENEURO.0352-19.2019.
URL https://www.eneuro.org/content/7/1/ENEURO.0352-19.2019

[33] S. L. Master, M. K. Eckstein, N. Gotlieb, R. Dahl, L. Wilbrecht, A. G. E. Collins, Disentangling the systems contributing to changes in learning during adolescence, Developmental Cognitive Neuroscience 41 (2020) 100732. doi:10.1016/j.dcn.2019.100732.
URL http://www.sciencedirect.com/science/article/pii/S1878929319303196

[34] M. K. Eckstein, S. L. Master, R. E. Dahl, L. Wilbrecht, A. G. E. Collins, Understanding the Unique Advantage of Adolescents in Stochastic, Volatile Environments: Combining Reinforcement Learning and Bayesian Inference, bioRxiv (2020) 2020.07.04.187971Publisher: Cold Spring Harbor Laboratory Section: New Results. doi:10.1101/2020.07.04.187971.
URL https://www.biorxiv.org/content/10.1101/2020.07.04.187971v1

[35] L. Xia, S. L. Master, M. K. Eckstein, B. Baribault, R. E. Dahl, L. Wilbrecht, A. G. E. Collins, Modeling changes in probabilistic reinforcement learning during adolescence, PLOS Computational Biology 17 (7) (2021) e1008524, publisher: Public Library of Science. doi:10.1371/journal.pcbi.1008524.
URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1

[36] S. J. Gershman, Empirical priors for reinforcement learning models, Journal of Mathematical Psychology 71 (2016) 1–6. doi:10.1016/j.jmp.2016.01.006.
URL http://www.sciencedirect.com/science/article/pii/S0022249616000080

[37] T. Harada, Learning From Success or Failure? – Positivity Biases Revisited, Frontiers in Psychology 11 (Jul. 2020). doi:10.3389/fpsyg.2020.01627.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7396482/

[38] Z. A. Yaple, R. Yu, Fractionating adaptive learning: A meta-analysis of the reversal learning paradigm, Neuroscience & Biobehavioral Reviews 102 (2019) 85–94. doi:10.1016/j.neubiorev.2019.04.006.
URL http://www.sciencedirect.com/science/article/pii/S0149763418308996

[39] X. Liu, J. Hairston, M. Schrier, J. Fan, Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies, Neuroscience and Biobehavioral Reviews 35 (5) (2011) 1219–1236. doi:10.1016/j.neubiorev.2010.12.012.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3395003/

[40] W. Schultz, A. Dickinson, Neuronal Coding of Prediction Errors, Annual Review of Neuroscience 23 (1) (2000) 473–500. doi:10.1146/annurev.neuro.23.1.473.
URL http://www.annualreviews.org/doi/10.1146/annurev.neuro.23.1.473

[41] A. G. E. Collins, J. K. Brown, J. M. Gold, J. A. Waltz, M. J. Frank, Working Memory Contributions to Reinforcement Learning Impairments in Schizophrenia, Journal of Neuroscience 34 (41) (2014) 13747–13756, publisher: Society for Neuroscience Section: Articles. doi:10.1523/JNEUROSCI.0989-14.2014.
URL https://www.jneurosci.org/content/34/41/13747

[42] A. Heinz, L. Deserno, U. S. Zimmermann, M. N. Smolka, A. Beck, F. Schlagenhauf, Targeted intervention: Computational approaches to elucidate and predict relapse in alcoholism, NeuroImage 151 (2017) 33–44. doi:10.1016/j.neuroimage.2016.07.055.
URL https://linkinghub.elsevier.com/retrieve/pii/S105381191630369X

[43] M. R. Nassar, M. J. Frank, Taming the beast: extracting generalizable knowledge from computational models of cognition, Current Opinion in Behavioral Sciences 11 (2016) 49–54. doi:10.1016/j.cobeha.2016.04.003.
URL http://linkinghub.elsevier.com/retrieve/pii/S2352154616300808

[44] A. H. Javadi, D. H. K. Schmidt, M. N. Smolka, Adolescents adapt more slowly than adults to varying reward contingencies, Journal of Cognitive Neuroscience 26 (12) (2014) 2670–2681.

[45] S.-J. Blakemore, T. W. Robbins, Decision-making in the adolescent brain, Nature Neuroscience 15 (9) (2012) 1184–1191, number: 9 Publisher: Nature Publishing Group. doi:10.1038/nn.3177.
URL http://www.nature.com/articles/nn.3177

[46] S. DePasque, A. Galván, Frontostriatal development and probabilistic reinforcement learning during adolescence, Neurobiology of Learning and Memory 143 (2017) 1–7. doi:10.1016/j.nlm.2017.04.009.
URL http://www.sciencedirect.com/science/article/pii/S107474271730062X

[47] A. Mohebi, J. R. Pettibone, A. A. Hamid, J.-M. T. Wong, L. T. Vinson, T. Patriarchi, L. Tian, R. T. Kennedy, J. D. Berke, Dissociable dopamine dynamics for learning and motivation, Nature 570 (7759) (2019) 65–70, number: 7759 Publisher: Nature Publishing Group. doi:10.1038/s41586-019-1235-y.
URL https://www.nature.com/articles/s41586-019-1235-y

[48] J. Davidow, K. Foerde, A. Galvan, D. Shohamy, An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence, Neuron 92 (1) (2016) 93–99. doi:10.1016/j.neuron.2016.08.031.
URL http://linkinghub.elsevier.com/retrieve/pii/S0896627316305244

[49] J. H. Decker, F. S. Lourenco, B. B. Doll, C. A. Hartley, Experiential reward learning outweighs instruction prior to adulthood, Cog-

nitive, Affective & Behavioral Neuroscience 15 (2) (2015) 310–320. doi:10.3758/s13415-014-0332-5.

[50] S. Palminteri, E. J. Kilford, G. Coricelli, S.-J. Blakemore, The Computational Development of Reinforcement Learning during Adolescence, PLoS Computational Biology 12 (6) (Jun. 2016). doi:10.1371/journal.pcbi.1004953.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4920542/

[51] W. R. Uttal, On some two-way barriers between models and mechanisms, Perception & Psychophysics 48 (2) (1990) 188–203. doi:10.3758/BF03207086.
URL https://doi.org/10.3758/BF03207086

[52] B. Webb, Can robots make good models of biological behaviour?, Behavioral and Brain Sciences 24 (6) (2001) 1033–1050, publisher: Cambridge University Press. doi:10.1017/S0140525X01000127.
URL http://www.cambridge.org/core/journals/behavioral-and-brain-sciences/arti

[53] D. J. Navarro, Between the Devil and the Deep Blue Sea: Tensions Between Scientific Judgement and Statistical Model Selection, Computational Brain & Behavior 2 (1) (2019) 28–34. doi:10.1007/s42113-018-0019-z.
URL https://doi.org/10.1007/s42113-018-0019-z

[54] T. Yarkoni, The generalizability crisis, The Behavioral and brain sciencesPublisher: Behav Brain Sci (Dec. 2020). doi:10.1017/S0140525X20001685.
URL https://pubmed.ncbi.nlm.nih.gov/33342451/

[55] T. E. J. Behrens, M. W. Woolrich, M. E. Walton, M. F. S. Rushworth, Learning the value of information in an uncertain world, Nature Neuroscience 10 (9) (2007) 1214–1221. doi:10.1038/nn1954.
URL https://www.nature.com/articles/nn1954

[56] J. McGuire, M. Nassar, J. Gold, J. Kable, Functionally Dissociable Influences on Learning Rate in a Dynamic Environment, Neuron 84 (4) (2014) 870–881. doi:10.1016/j.neuron.2014.10.013.
URL https://linkinghub.elsevier.com/retrieve/pii/S0896627314009118

[57] M. D. Lee, How cognitive modeling can benefit from hierarchical Bayesian models, Journal of Mathematical Psychology 55 (1) (2011) 1–7. doi:10.1016/j.jmp.2010.08.013.
URL https://linkinghub.elsevier.com/retrieve/pii/S0022249610001148

[58] A. G. E. Collins, M. J. Frank, How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis: Working memory in reinforcement learning, European Journal of Neuroscience 35 (7) (2012) 1024–1035. doi:10.1111/j.1460-9568.2011.07980.x.

[59] L. H. Somerville, S. F. Sasse, M. C. Garrad, A. T. Drysdale, N. Abi Akar, C. Insel, R. C. Wilson, Charting the expansion of strategic exploratory behavior during adolescence, Journal of Experimental Psychology: General 146 (2) (2017) 155–164, place: US Publisher: American Psychological Association. doi:10.1037/xge0000250.

[60] A. Gopnik, Childhood as a solution to explore–exploit tensions, Philosophical Transactions of the Royal Society B: Biological Sciences 375 (1803) (2020) 20190502, publisher: Royal Society. doi:10.1098/rstb.2019.0502.
URL https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0502

[61] V. M. Brown, J. Chen, C. M. Gillan, R. B. Price, Improving the Reliability of Computational Analyses: Model-Based Planning and Its Relationship With Compulsivity, Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 5 (6) (2020) 601–609. doi:10.1016/j.bpsc.2019.12.019.
URL https://www.sciencedirect.com/science/article/pii/S2451902220300161

[62] I. W. Eisenberg, P. G. Bissett, A. Zeynep Enkavi, J. Li, D. P. MacKinnon, L. A. Marsch, R. A. Poldrack, Uncovering the structure of self-regulation through data-driven ontology discovery, Nature Communications 10 (1) (2019) 1–13. doi:10.1038/s41467-019-10301-1.
URL https://www.nature.com/articles/s41467-019-10301-1

[63] W. C. Lin, K. Delevich, L. Wilbrecht, A role for adaptive developmental plasticity in learning and decision making, Current Opinion in Behavioral Sciences 36 (2020) 48–54. doi:10.1016/j.cobeha.2020.07.010.
URL http://www.sciencedirect.com/science/article/pii/S2352154620301121

[64] M. R. Nassar, R. Bruckner, J. I. Gold, S.-C. Li, H. R. Heekeren, B. Eppinger, Age differences in learning emerge from an insufficient representation of uncertainty in older adults, Nature Communications 7 (2016) 11609. doi:10.1038/ncomms11609.
URL http://www.nature.com/doifinder/10.1038/ncomms11609

[65] K. C. Berridge, The debate over dopamine's role in reward: the case for incentive salience, Psychopharmacology 191 (3) (2007) 391–431. doi:10.1007/s00213-006-0578-x.
URL https://doi.org/10.1007/s00213-006-0578-x

[66] A. J. Yu, P. Dayan, Uncertainty, Neuromodulation, and Attention, Neuron 46 (4) (2005) 681–692. doi:10.1016/j.neuron.2005.04.026.
URL http://www.sciencedirect.com/science/article/pii/S0896627305003624

[67] S. Bouret, S. J. Sara, Network reset: a simplified overarching theory of locus coeruleus noradrenaline function, Trends in Neurosciences 28 (11) (2005) 574–582. doi:10.1016/j.tins.2005.09.002.

[68] S. J. Gershman, Dopamine, Inference, and Uncertainty, Neural Computation 29 (12) (2017) 3311–3326. doi:10.1162/neco$_{a0}$1023.
URLhttp://www.mitpressjournals.org/doi/abs/10.1162/neco$_{a0}$1023$C.\ K.Starkweath$

[69] S. J. Gershman, N. Uchida, Believing in dopamine, Nature Reviews Neuroscience 20 (11) (2019) 703–714, number: 11 Publisher: Nature Publishing Group. doi:10.1038/s41583-019-0220-7.
URL https://www.nature.com/articles/s41583-019-0220-7

[70] M. Jepma, P. R. Murphy, M. R. Nassar, M. Rangel-Gomez, M. Meeter, S. Nieuwenhuis, Catecholaminergic Regulation of Learning Rate in a Dynamic Environment, PLOS Computational Biology 12 (10) (2016) e1005171, publisher: Public Library of Science. doi:10.1371/journal.pcbi.1005171.
URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.100517

[71] K. Katahira, The statistical structures of reinforcement learning with asymmetric value updates, Journal of Mathematical Psychology 87 (2018) 31–45. doi:10.1016/j.jmp.2018.09.002.
URL http://www.sciencedirect.com/science/article/pii/S0022249617302407

[73] M. Sugawara, K. Katahira, Dissociation between asymmetric value updating and perseverance in human reinforcement learning, Scientific Reports 11 (1) (2021) 3574, number: 1 Publisher: Nature Publishing Group. doi:10.1038/s41598-020-80593-7.
URL https://www.nature.com/articles/s41598-020-80593-7

[74] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, S. J. Gershman, Building machines that learn and think like people, Behavioral and Brain Sciences 40 (2017). doi:10.1017/S0140525X16001837.
URL https://www.cambridge.org/core/product/identifier/S0140525X16001837/type/jour

[75] A. G. E. Collins, Reinforcement learning: bringing together computation and cognition, Current Opinion in Behavioral Sciences 29 (2019) 63–68. doi:10.1016/j.cobeha.2019.04.011.
URL http://www.sciencedirect.com/science/article/pii/S235215461830175X

[76] A. G. E. Collins, The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory, Journal of Cognitive Neuroscience 30 (10) (2018) 1422–1432. doi:10.1162/jocn$_{a0}$1238.

[77] L.-H. Tai, A. M. Lee, N. Benavidez, A. Bonci, L. Wilbrecht, Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value, Nature Neuroscience 15 (9) (2012) 1281–1289. doi:10.1038/nn.3188.

[78] M. Donoso, A. G. E. Collins, E. Koechlin, Foundations of human reasoning in the prefrontal cortex, Science 344 (6191) (2014) 1481–1486. doi:10.1126/science.1252254.
URL http://www.sciencemag.org/cgi/doi/10.1126/science.1252254

[79] A. M. Bornstein, K. A. Norman, Reinstated episodic context guides sampling-based decisions for reward, Nature Neuroscience 20 (7) (2017) 997–1003. doi:10.1038/nn.4573.
URL https://www.nature.com/articles/nn.4573

[80] O. M. Vikbladh, M. R. Meager, J. King, K. Blackmon, O. Devinsky, D. Shohamy, N. Burgess, N. D. Daw, Hippocampal Contributions to Model-Based Planning and Spatial Memory, Neuron 102 (3) (2019) 683–693.e4. doi:10.1016/j.neuron.2019.02.014.
URL https://www.sciencedirect.com/science/article/pii/S0896627319301230

[81] M. E. van der Schaaf, E. Warmerdam, E. A. Crone, R. Cools, Distinct linear and non-linear trajectories of reward and punishment reversal learning during development: relevance for dopamine's role in adolescent decision making, Developmental Cognitive Neuroscience 1 (4) (2011) 578–590. doi:10.1016/j.dcn.2011.06.007.

[82] N. Sendhilnathan, M. Semework, M. E. Goldberg, A. E. Ipata, Neural Correlates of Reinforcement Learning in Mid-lateral Cerebellum, Neuron 106 (1) (2020) 188–198.e5. doi:10.1016/j.neuron.2019.12.032.

[83] S. D. McDougle, A. G. E. Collins, Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning, Psychonomic Bulletin & Review 28 (1) (2021) 20–39. doi:10.3758/s13423-020-01774-z.
URL https://doi.org/10.3758/s13423-020-01774-z

[84] A. Radulescu, Y. Niv, I. Ballard, Holistic Reinforcement Learning: The Role of Structure and Attention, Trends in Cognitive Sciences 23 (4) (2019) 278–292. doi:10.1016/j.tics.2019.01.010.
URL https://www.sciencedirect.com/science/article/pii/S1364661319300361

[85] A. Konovalov, I. Krajbich, Neurocomputational Dynamics of Sequence Learning, Neuron 98 (6) (2018) 1282–1293.e4. doi:10.1016/j.neuron.2018.05.013.
URL http://www.sciencedirect.com/science/article/pii/S0896627318303854

[86] W. Kool, F. A. Cushman, S. J. Gershman, When Does Model-Based Control Pay Off?, PLOS Computational Biology 12 (8) (2016) e1005090, publisher: Public Library of Science. doi:10.1371/journal.pcbi.1005090.
URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.100509

[87] C. F. d. Silva, T. A. Hare, Humans are primarily model-based learners in the two-stage task, bioRxiv (2020) 682922Publisher: Cold Spring Harbor Laboratory Section: New Results. doi:10.1101/682922.
URL https://www.biorxiv.org/content/10.1101/682922v4

[88] C. Johnson, L. Wilbrecht, Juvenile mice show greater flexibility in multiple choice reversal learning than adults, Developmental Cognitive Neuroscience 1 (4) (2011) 540–551. doi:10.1016/j.dcn.2011.05.008.
URL http://linkinghub.elsevier.com/retrieve/pii/S1878929311000533

[89] A. C. Petersen, L. Crockett, M. Richards, A. Boxer, A self-report measure of pubertal status: Reliability, validity, and initial norms, Journal of Youth and Adolescence 17 (2) (1988) 117–133. doi:10.1007/BF01537962.
URL https://doi.org/10.1007/BF01537962

[90] A. Christakou, S. J. Gershman, Y. Niv, A. Simmons, M. Brammer, K. Rubia, Neural and psychological maturation of decision-making in adolescence and young adulthood, Journal of Cognitive Neuroscience 25 (11) (2013) 1807–1823.

[91] W. van den Bos, M. X. Cohen, T. Kahnt, E. A. Crone, Striatum–Medial Prefrontal Cortex Connectivity Predicts Developmental Changes in Reinforcement Learning, Cerebral Cortex 22 (6) (2012) 1247–1255. doi:10.1093/cercor/bhr198.
URL https://academic.oup.com/cercor/article/22/6/1247/307075

[92] M. J. Frank, L. C. Seeberger, R. C. O'Reilly, By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism, Science 306 (5703) (2004) 1940–1943. doi:10.1126/science.1102941.

[93] R. D. Cazé, M. A. A. van der Meer, Adaptive properties of differential learning rates for positive and negative outcomes, Biological Cybernetics 107 (6) (2013) 711–719. doi:10.1007/s00422-013-0571-5.

[94] G. Lefebvre, M. Lebreton, F. Meyniel, S. Bourgeois-Gironde, S. Palminteri, Behavioural and neural characterization of optimistic reinforcement learning, Nature Human Behaviour 1 (4) (2017) 0067. doi:10.1038/s41562-017-0067.
URL http://www.nature.com/articles/s41562-017-0067

[95] W. Dabney, Z. Kurth-Nelson, N. Uchida, C. K. Starkweather, D. Hassabis, R. Munos, M. Botvinick, A distributional code for value in dopamine-based reinforcement learning, Nature 577 (7792) (2020) 671–675. doi:10.1038/s41586-019-1924-6.
URL http://www.nature.com/articles/s41586-019-1924-6

[96] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian Data Analysis, 3rd Edition, Chapman and Hall/CRC, Boca Raton, 2013.

[97] K. Katahira, How hierarchical models improve point estimates of model parameters at the individual level, Journal of Mathematical Psychology 73 (2016) 37–58. doi:10.1016/j.jmp.2016.03.007.

[98] S. Watanabe, A Widely Applicable Bayesian Information Criterion, Journal of Machine Learning Research 14 (Mar) (2013) 867–897.
URL http://www.jmlr.org/papers/v14/watanabe13a.html

[99] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12 (85) (2011) 2825–2830.
URL http://jmlr.org/papers/v12/pedregosa11a.html

[100] D. A. Peterson, C. Elliott, D. D. Song, S. Makeig, T. J. Sejnowski, H. Poizner, Probabilistic reversal learning is impaired in Parkinson's disease, Neuroscience 163 (4) (2009) 1092–1101. doi:10.1016/j.neuroscience.2009.07.033.
URL http://www.sciencedirect.com/science/article/pii/S0306452209012068

[101] R. Swainson, R. D. Rogers, B. J. Sahakian, B. A. Summers, C. E. Polkey, T. W. Robbins, Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: possible adverse effects of dopaminergic medication, Neuropsychologia 38 (5) (2000) 596–612. doi:10.1016/S0028-3932(99)00103-7.
URL http://www.sciencedirect.com/science/article/pii/S0028393299001037

[102] J. A. Waltz, J. M. Gold, Probabilistic reversal learning impairments in schizophrenia: Further evidence of orbitofrontal dysfunction, Schizophrenia Research 93 (1) (2007) 296–303. doi:10.1016/j.schres.2007.03.010.
URL http://www.sciencedirect.com/science/article/pii/S092099640700120X

[103] D. P. Dickstein, E. C. Finger, M. A. Brotman, B. A. Rich, D. S. Pine, J. R. Blair, E. Leibenluft, Impaired probabilistic reversal learning in youths with mood and anxiety disorders, Psychological Medicine 40 (7) (2010) 1089–1100. doi:10.1017/S0033291709991462.
URL http://www.cambridge.org/core/journals/psychological-medicine/article/impaire

[104] R. Cools, L. Clark, A. M. Owen, T. W. Robbins, Defining the Neural Mechanisms of Probabilistic Reversal Learning Using Event-Related Functional

Magnetic Resonance Imaging, Journal of Neuroscience 22 (11) (2002) 4563–4567. doi:10.1523/JNEUROSCI.22-11-04563.2002.
URL https://www.jneurosci.org/content/22/11/4563

[105] R. Cools, M. J. Frank, S. E. Gibbs, A. Miyakawa, W. Jagust, M. D'Esposito, Striatal Dopamine Predicts Outcome-Specific Reversal Learning and Its Sensitivity to Dopaminergic Drug Administration, Journal of Neuroscience 29 (5) (2009) 1538–1543. doi:10.1523/JNEUROSCI.4467-08.2009.
URL https://www.jneurosci.org/content/29/5/1538

[106] F. Lourenco, B. Casey, Adjusting behavior to changing environmental demands with development, Neuroscience & Biobehavioral Reviews 37 (9) (2013) 2233–2242. doi:10.1016/j.neubiorev.2013.03.003.
URL https://linkinghub.elsevier.com/retrieve/pii/S0149763413000638

[107] A. Izquierdo, J. L. Brigman, A. K. Radke, P. H. Rudebeck, A. Holmes, The neural basis of reversal learning: An updated perspective, Neuroscience 345 (2017) 12–26. doi:10.1016/j.neuroscience.2016.03.021.
URL http://www.sciencedirect.com/science/article/pii/S030645221600244X

[108] A. G. E. Collins, B. Ciullo, M. J. Frank, D. Badre, Working Memory Load Strengthens Reward Prediction Errors, The Journal of Neuroscience 37 (16) (2017) 4332–4342. doi:10.1523/JNEUROSCI.2700-16.2017.
URL http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2700-16.2017

[109] A. G. E. Collins, M. A. Albrecht, J. A. Waltz, J. M. Gold, M. J. Frank, Interactions Among Working Memory, Reinforcement Learning, and Effort in Value-Based Choice: A New Paradigm and Selective Deficits in Schizophrenia, Biological Psychiatry 82 (6) (2017) 431–439. doi:10.1016/j.biopsych.2017.05.017.
URL http://www.sciencedirect.com/science/article/pii/S0006322317316190

[110] A. G. E. Collins, M. J. Frank, Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory, Proceedings of the National Academy of Sciences 115 (10) (2018) 2502–2507, publisher: National Academy of Sciences Section: Biological Sciences. doi:10.1073/pnas.1720963115.
URL https://www.pnas.org/content/115/10/2502

Appendix A. Behavioral Measures

In addition to RL model parameters, we also extracted behavioral features from each of the three tasks, conducted regression analyses, and in one case fitted a non-RL computational model. The resulting features of these analyses were used in the PCA (2.2.1; 4.6; Fig. 3; suppl. Fig. D.10), behavioral Ridge regression (2.2.3; 4.7), and comprehensive correlation matrix (suppl. Fig. D.9). The following list details these features, using the labels in Fig. 3, suppl. Fig. D.10, and suppl. Fig. D.9 to denote each feature:

- Basic performance measures
  - "Missed trials" (percentage of missed trials)
  - "Response times" (average response times on correct trials)
  - "Response time variability" (standard deviation of response times on correct trials)
  - "Percent correct" (accuracy; overall percentage of correct trials)
- Advanced performance measures
  - "Acc. first 3 trials" (only for task A; average accuracy on the first three trials of each stimulus; a measure of initial learning)
  - "Asymptotic acc." (only for task A; an exponential curve was fit to the learning curve of each subject. The functional form was: $0.5+a-a*exp(-c*(t-1))$, where t is trial number. a is bounded $0 < a < 0.5$, and $c > 0$. This feature refers to the asymptote of the curve: $0.5+a-a*exp(-c*(T-1))$, where T is total number of trials completed by the subject)
  - "Learning slope" (only for task A; this feature refers to the slope $log(c)$ of the model above)
  - "Acc. intercept" (only for task A; hierarchical regression was fit to predict the probability of a correct choice on each trial. The regression formula was $p(correct)\ 1+prew+delay+(1+prew+delay|id)$, where $(...|id)$ indicates random effects by subject id. This feature refers to the random intercept for each subject)
  - "Reward effect on acc." (in the same regression model, this feature refers to the random slope for the reward history predictor, which is the z-scored number of correct trials for each butterfly)

- "Delay" (only tasks A and C; in the same [or an equivalent] regression model, this feature refers to the random slope for the delay predictor, which is the z-scored number of trials since last time the same stimulus was seen and the participant chose the correct action)
- "Number of switches" (only for task B; the number of task switches experienced by each participant; because the occurrence of switches was based on performance, the number of switches is an additional criterion of task performance)
- "Criterion after switch" (only for task B; number of trials after a task switch until participants reached a performance criterion of 2 correct choices after the task switch)
- "Acc. late minus early" (only for task B; difference in accuracy between the first and last third of trials in the task, as a measure of "slow" learning)

• Action repetition

- "Stay (choice)" (for tasks A and C; percentage of choices that were repeated with respect to each stimulus, averaged across stimuli)
- "Stay (motor)" (for tasks A and C; percentage of choices that were repeated between two subsequent trials, irrespective of the shown stimulus)
- "Stay (both)" (for task B; because the task only provided a single stimulus, stay-choice and stay-motor were identical)
- "Win-stay" (Fraction of trials in which participants repeated a choice for a given stimulus that was rewarded on the previous trial for that stimulus, normalized by number of win trials, and averaged over stimuli: $\frac{win\ stay}{win\ stay + win\ shift}$)
- "Win-lose-stay" (for task B only; fraction of trials in which participants repeated a choice that was rewarded two trials back, but not rewarded on the previous trial, normalized by number of win-lose trials)
- "Lose-stay" (Fraction of trials in which participants repeated a choice for a given stimulus that was not rewarded on the previous

trial for that stimulus, normalized by number of lose trials, and averaged over stimuli: $\frac{lose\ stay}{lose\ stay + lose\ shift}$)

- Bayesian model parameters (task B only)
  - "$p(switch)$ param." (a Bayesian inference model was fit in addition to the RL model. The Bayesian model employed a mental model of the task, which was based on two hidden states "Left is correct" and "Right is correct", and used Bayesian inference to infer the current hidden state based on recent outcomes. The free parameters of this model were the task parameters of the mental model, switch probability on each trial p_{switch}, and probability of reward for a correct choice p_{reward}, choice parameters Persistence P and inverse decision temperature β. Detailed information about this model is provided in [34]. This feature refers to model parameter p_{switch}.)
  - "$p(reward)$ param." (parameter $p(switch)$ in the Bayesian model)
- Behavioral measures of working memory (task C only)
  - "Set size effect on acc." (a logistic regression was fitted to choice data, assessing the effects of set size, delay, number of previous correct trials, and number of previous incorrect trials on choice accuracy; this feature refers to the effect of set size)
  - "Feature $p(cor)$" (this features refers to the effect of the number of previous correct trials)
  - "Feature $p(inc)$" (this features refers to the effect of the number of previous incorrect trials)
  - "Set size Acc. slope" (performance was averaged for blocks of each set size, and the slope in performance over set sizes was determined)
  - "Set size Acc. slope" (similar to the previous feature, but replacing performance with response times)

Appendix B. Previous Results

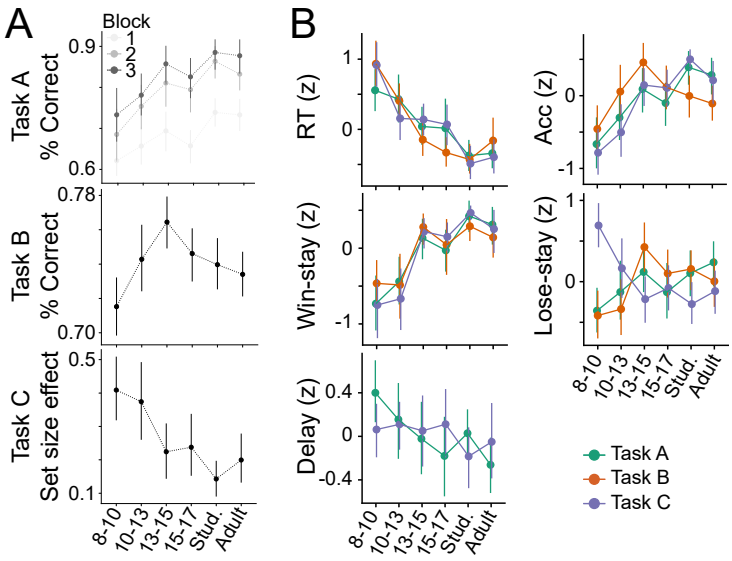

Figure B.6: Main results of tasks A, B, and C. (A) Top: In task A, performance increased with age and plateaued in early adulthood, as captured in decreases in decision temperature $\frac{1}{\beta}$ and increases in learning rate α [35]. Performance also increased over task time (blocks). Middle: In task B, performance showed a remarkable inverse U-shaped age trajectory: Performance increased markedly from early childhood (8-10 years) to mid-adolescence (13-15), but decreased in late adolescence (15-17) and adulthood (18-30; [34]). Bottom: Task C showed that the effect of set size on performance (regression coefficient) decreased with age, which was captured by increases in RL learning rate, but stable WM limitations [33]. (B) Main behavioral features over age; colors denote task; all features are z-scored. Some measures (e.g., response times [RT], win-stay choices) were consistent across tasks, while others (e.g., accuracy [Acc.], lose-stay choices) showed significant differences (see Table C.7).

Each task was first analyzed independently, and detailed results have been presented elsewhere [33, 35, 34]. We summarize the key results here. In task A, participants saw one of four butterflies on each trial, and aimed to pick the one of two flowers that was preferred by this butterfly. Each butterfly had a stable preference for one flower, and participants received a stochastic reward (80% probability) when they chose this flower. Nevertheless, sometimes the butterfly liked the opposite flower, and participants got a reward with 20% when they chose the opposite flower (Fig. 1C; 4.3). Task A has

been used previously to investigate the role of reward sensitivity and its interplay with episodic memory, shedding light on the neural substrate of these processes, notably the striatum and hippocampus, and revealing a unique role of adolescence in stochastic learning [48]. In our sample, performance on task A increased with age through the early-twenties and then stabilized [35]; suppl. Fig. B.6A). Using hierarchical Bayesian methods to fit RL models, we showed that this performance increase was driven by increasing positive learning rate α_+ and decreasing decision noise $\frac{1}{\beta}$. Forgetting rates decreased very slightly with age, and negative learning rate α_- was 0, suggesting that participants ignored negative outcomes (Fig. 2A and 2C).

In task B, participants saw two boxes and selected one on each trial, with the goal of collecting gold coins. For some period of time, one box was correct and led to a stochastic reward (75% probability), while the other was unrewarded (0% probability). Then, the contingencies switched unpredictably and unsignaled, and the opposite box became the correct one. A 120-trials session contained 2-7 switches (Fig. 1D; 4.3). Task B was adapted from the rodent literature, where it has been used to show a causal link between stimulation of striatal spiny projection neurons and subsequent choices [77]. Stochastic reversal tasks are also common in the human literature [100, 101, 81, 102, 103, 104, 105, 106, 107]. A notable feature of our task compared to others is the deterministic feedback for incorrect choices. In our study, we found that human youth age 13-15 years markedly outperformed younger youth (8-12), older youth (16-17), and even young adults (18-30), suggesting that adolescent brains might be specifically adapted to perform well in the stochastic and volatile environment of task B. Computational modeling, using hierarchical Bayesian fitting, revealed that some model parameters (e.g,. decision temperature $\frac{1}{\beta}$, Persistence) increased monotonically from childhood to adulthood, whereas others (e.g., learning rate for negative feedback α_-, Bayesian inference parameters p_{switch} and p_{reward}) showed pronounced U-shapes with peaks in 13-to-15-year-olds, similar to performance. Blending RL and a Bayesian inference models using PCA revealed that adolescents operated at a sweet spot that combined mature levels of task performance with child-like, short time scales of learning, and provided an explanation for adolescents' superior performance [34].

Task C was designed to dissociate the effects of RL and working memory, and has been used in diverse samples of adult participants [58, 108, 41, 109, 110, 76, 83], but this study was the first to test it in a developmental samples

[33]. In this task, participants saw one stimulus at a time (e.g., bee) and chose one of three actions in response (left, up, right; Fig. 1E, right). Feedback was deterministic, i.e., reliably identified each action as correct or incorrect. The goal of task C was to learn the correct response for each stimulus. The key feature of the task is that stimuli appear in independent blocks of different sizes, ranging from 2-5 stimuli (e.g., the bee could be presented in a block containing just 1 other animal, or up to 4 other animals). As set sizes increase, participants have been shown to shift the balance between using their capacity-limited, but reliable working memory system, to using their unlimited, but slower RL system [58]. Task C estimates both memory systems, RL as well as working memory. We found that participants aged 8-12 learned slower than participants aged 13-17, and were more sensitive to set size (suppl. Fig. B.6A). Computational modeling revealed that developmental changes in RL were more protracted than changes in working memory: RL learning rate α_+ increased until age 18, whereas WM parameters showed weaker and more subtle changes early in adolescence [33].

Appendix C. Age Trajectories of Behavioral Measures

This section provides additional information about selected behavioral features (section Appendix A) and assesses their development with age. For statistical testing, we assessed age trajectories using similar regression models as before (section 2.1.3).

Response times, reflecting choice fluidity and task engagement, sped up with age in all tasks, whereby age trajectories differed significantly between tasks A and B in pairwise follow-up models (grand model $AIC_{with\ task} = 1.868$, $AIC_{no\ task} = 1.871$; for detailed statistics, see Table C.7; suppl. Fig. B.6B). Accuracy, reflecting subjective ease and task engagement, showed a significant increase with age, and no significant pairwise differences in age trajectories after correcting for multiple comparisons, despite the better fit of the model including task compared to the model without task ($AIC_{with\ task} = 2.015$, $AIC_{no\ task} = 2.024$; suppl. Fig. B.6B; Table C.7).

Win-stay (WS) behavior reflects participants' tendency to repeat rewarded actions, while lose-stay (LS) behavior reflects participants' tendency to repeat non-rewarded actions. Win-stay behavior increased with age, without task differences ($AIC_{with\ task} = 1.961$, $AIC_{no\ task} = 1.959$; suppl. Fig. B.6B; Table C.7). Lose-stay behavior showed marked task differences ($AIC_{with\ task} = 2.075$, $AIC_{no\ task} = 2.109$), with inverse trajectories in task C compared to

the other tasks: In task C, lose-stay behavior decreased monotonically until mid-adolescence (linear effect of age: $w = -0.31$, $p < 0.001$; quadratic effect: $w = 0.007$, $p < 0.001$), whereas in task A, it increased slightly (linear effect of age: $w = 0.075$, $p < 0.001$; quadratic effect: $w = -0.001$, $p < 0.001$). In task B, lose-stay behavior showed an inverse-U trajectory (linear effect: $w = 0.20$, $p < 0.001$; quadratic: $w = -0.005$, $p < 0.001$; suppl. Fig. B.6B). These differences mirrored differences in optimal task strategies: Lose-stay is always a bad strategy in task C because negative feedback is diagnostic and actions with negative outcomes should never be repeated. In tasks A and B, on the other hand, some proportion of lose-stay choices is necessary because individual negative feedback is not diagnostic, and several pieces of evidence need to be integrated over time.

Lastly, the Delay pattern measured the decrease in accuracy with increasing delay between two presentations of the same stimulus (section Appendix A). Delay did not show significant age changes (Table C.7), and did not differ between tasks ($AIC_{with\ task} = 1.405$, $AIC_{no\ task} = 1.402$).

Appendix D. Other Supplemental Figures

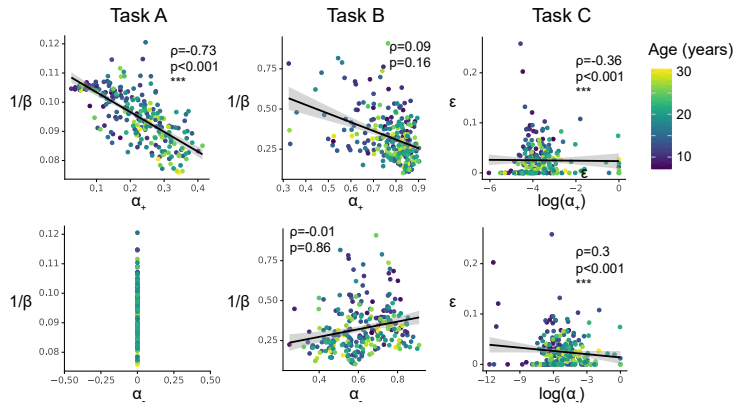

Figure D.7: Within-task parameter correlations, focusing on learning rates (x-axes) and exploration / noise parameters (y-axes). Each column shows one task. Each dot in the scatter plots refers to a participant, colors indicates age. Inserted are Spearman correlation statistics.

Table C.7: Statistics of mixed-effects regression models predicting z-scored behavioral features from task (task A, task B, task C), age, and squared age (months). The task-less grand model is reported when it had the best model fit (win-stay, Delay). Otherwise, pairwise follow-up models are shown (RT, ACC, lose-stay), with p-values corrected for multiple comparison using the Bonferroni correction. * $p < .05$; ** $p < .01$, *** $p < .001$.

Parameter	Tasks	Predictor	β	p (Bonf.)	sig.
RT	task B & task A	Task (main effect)	2.15	0.006	**
		Linear age	-0.23	0.003	**
		Task * linear age (interaction)	-0.25	0.003	**
		Task * quadratic age (interaction)	0.007	0.003	**
	task B & task C	Task (main effect)	-0.76	0.69	
		Linear age	-0.48	< 0.001	***
		Task * linear age (interaction)	0.10	0.45	
		Task * quadratic age (interaction)	-0.003	0.288	
	task A & task C	Task (main effect)	1.40	0.63	
		Linear age	-0.23	0.003	**
		Task * linear age (interaction)	-0.15	0.084	
		Task * quadratic age (interaction)	0.004	0.129	
ACC	task B & task A	Task (main effect)	1.27	0.36	
		Linear age	0.27	< 0.001	***
		Task * linear age (interaction)	-0.10	0.87	
		Task * quadratic age (interaction)	0.001	1	
	task B & task C	Task (main effect)	-2.15	0.033	*
		Linear age	0.17	0.036	*
		Task * linear age (interaction)	0.20	0.118	
		Task * quadratic age (interaction)	-0.004	0.33	
	task A & task C	Task (main effect)	-0.88	0.60	
		Linear age	0.27	< 0.001	***
		Task * linear age (interaction)	0.10	0.57	
		Task * quadratic age (interaction)	-0.003	0.57	
WS	—	Intercept	-3.05	< 0.001	***
		Age (linear)	0.31	< 0.001	***
		Age (quadratic)	-0.007	< 0.001	***
LS	task B & task A	Task (main effect)	-0.90	0.42	
		Linear age	0.075	0.87	
		Task * linear age (interaction)	0.12	0.42	
		Task * quadratic age (interaction)	-0.004	0.29	
	task B & task C	Task (main effect)	4.84	< 0.001	***
		Linear age	0.20	0.015	*
		Task * linear age (interaction)	-0.51	< 0.001	***
		Task * quadratic age (interaction)	0.012	< 0.001	***
	task A & task C	Task (main effect)	3.94	< 0.001	***
		Linear age	0.075	0.54	
		Task * linear age (interaction)	-0.39	< 0.001	***
		Task * quadratic age (interaction)	0.008	0.003	**
Delay	—	Intercept	0.95	0.035	*
		Age (linear)	-0.09	0.07	
		Age (quadratic)	0.002	0.14	

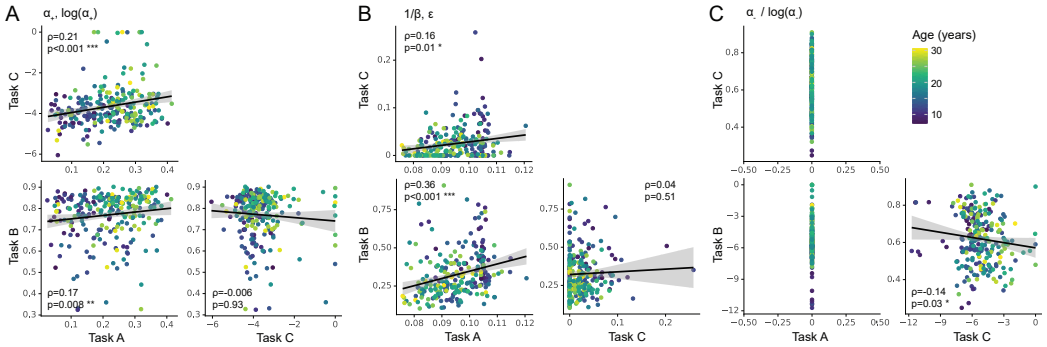

Figure D.8: Between-task parameter correlations. (A) Parameter α_+ across tasks ($log(\alpha_+$ in task C). (B) Parameters $\frac{1}{\beta}$ (task A, B) and ϵ (task C). (C) Parameter α_- across tasks ($log(\alpha_-)$ in task C). Same conventions as in Fig. D.7.

Figure D.9: Full Spearman correlation matrix of all features in the dataset. Feature order is the same as in Fig. 3. Deeper red (blue) colors indicate stronger positive (negative) correlations in terms of Spearman's ρ (see color legend). Only correlations with p-values < 0.05 are shown; remaining squares are left blank.

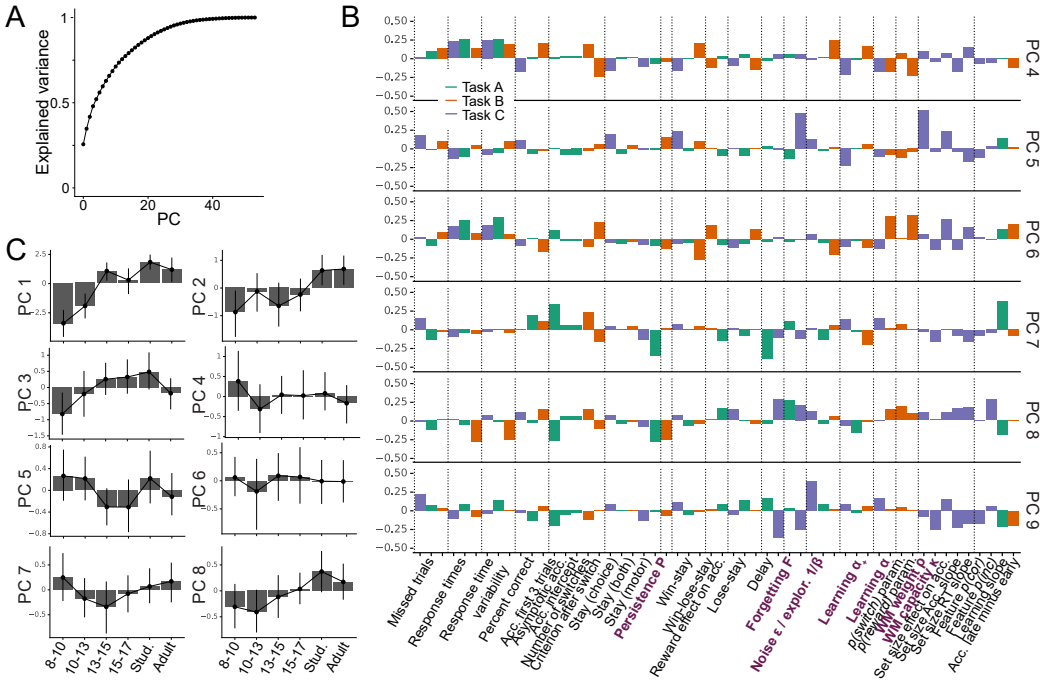

Figure D.10: Additional PCA results. (A) Cumulative variance explained by all PCs of the PCA (Fig. 3; 2.2.1). The smooth, non-stepped function does not provide evidence for lower-dimensional structure within the dataset. (B) Feature loadings (weights) of PC4-PC9. Loadings are flipped based on their relation to task performance, like for PC2-PC3 in Fig. 3. (C) Age trajectories of the top 8 PCs, by age group. Corresponding statistics in suppl. Table D.8.)

Table D.8: Statistics of regular regression models predicting each PC from two age predictors (linear and quadratic).

PC	Effect	β	t	p	sig.
1	age (linear)	1.56	6.56	< 0.001	***
	age (quadratic)	0.035	5.61	< 0.001	***
2	age (linear)	0.34	2.17	0.031	*
	age (quadratic)	0.007	1.64	0.10	—
3	age (linear)	0.46	3.27	0.001	**
	age (quadratic)	-0.011	-3.13	0.002	**

Table D.9: Statistics of mixed-effects regression models predicting parameter values from task (A, B, and C), age, and squared age (months). Only effects including task are reported. * $p < .05$; ** $p < .01$, *** $p < .001$.

Parameter	Tasks	Predictor	β	p	sig.
α_+	task B & task A	Task (main effect)	0.79	< 0.001	***
		Task * linear age (interaction)	-0.025	0.009	**
		Task * quadratic age (interaction)	0.001	0.021	*
	task B & task C	Task (main effect)	0.84	< 0.001	***
		Task * linear age (interaction)	-0.012	0.41	
		Task * quadratic age (interaction)	< 0.001	0.55	
	task A & task C	Task (main effect)	0.048	0.70	
		Task * linear age (interaction)	-0.12	0.37	
		Task * quadratic age (interaction)	< 0.001	0.36	
$\frac{1}{\beta}$	task B & task A	Task (main effect)	0.49	< 0.001	***
		Task * linear age (interaction)	-0.026	0.046	*
		Task * quadratic age (interaction)	0.001	< 0.001	***
α_-	task B & task C	Task (main effect)	11.70	< 0.001	***
		Task * linear age (interaction)	0.58	< 0.001	***
		Task * quadratic age (interaction)	-0.013	< 0.001	***
Forgetting	task B & task C	Task (main effect)	0.10	0.36	
		Task * linear age (interaction)	0.005	0.70	
		Task * quadratic age (interaction)	< 0.001	0.67	