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

- Efforts in computational cognitive modeling often assume that Reinforcement Learning (RL) modeling parameters will generalize between studies and models, but this is not well established.
- We empirically investigate whether RL parameters generalize between three tasks and models, using a large developmental dataset and a within-participant design.
- We find that RL decision noise/exploration parameters generalize fairly well, but RL learning rates do not.
- Our data support previous conclusions that decision noise/exploration decreases during development (ages 8-17), but suggests that claims about learning rate development cannot be generalized.

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) has revolutionized the cognitive and brain sciences, explaining behavior from simple conditioning to problem solving, across the life span, and anchored in brain function. However, discrepancies in results are increasingly apparent between studies, particularly in the developmental literature. To better understand these, we investigated to which extent parameters *generalize* between tasks and models, and capture specific and uniquely *interpretable* (neuro)cognitive processes. 291 participants aged 8-30 years completed three learning tasks in a single session, and were fitted using state-of-the-art RL models. RL decision noise/exploration parameters generalized well between tasks, decreasing between ages 8-17. Learning rates for negative feedback did not generalize, and learning rates for positive feedback showed intermediate generalizability, dependent on task similarity. These findings can explain discrepancies in the existing literature. Future research therefore needs to carefully consider task characteristics when relating findings across studies, and develop strategies to computationally model how context impacts behavior.

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 complex ones, 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 illnesses [14, 15, 16, 17, 18]. RL models are of particular interest because they also capture 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, not enough attention has been paid to what exactly RL models and model variables (e.g., model parameters) measure, and our current assumptions might be imprecise, potentially slowing further progress. Our recent opinion paper develops this argument in depth [28]. In brief, computational modeling condenses behavioral datasets into a model and a small number of free model parameters [11, 27, 29, 30, 31, 32]. We as researchers often assume that these models and parameters expose mental and/or neural processes, and have the ability to dissect them into specific, unique components (e.g., value updating and decision making), thereby measuring participants' inherent characteristics (e.g., individual learning rates). However, we argue in this paper that these assumptions might be too optimistic and that a careful empirical investigation is required to assess their validity.

We focus on two major aspects, which are adopted widely in computational modeling [28]: *generalizability* and *interpretability*. We define a model variable (e.g., fitted parameter) 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 (e.g., a person with a high learning rate) rather than task dependent. We further define a model parameter as *interpretable* if it isolates specific and unique elements of cognition, which are often assumed to be implemented in separable neural substrates: Decomposing behavior into model parameters is seen as a way of *carving cognition at its joints*.

Assumptions about generalizability and interpretability are rarely stated explicitly, but underlie conclusions across the fields of computational psychology and neuroscience, and often implicitly guide research efforts. Assumptions of generalizability, for example, inspired many to identify the inherent, task-independent settings of parameters in humans (e.g., empirical parameter distributions [33]; relationships between negative and positive learning rates [34]), to characterize the age development of parameters in a task-independent way [11, 12, 13, 35], and to compare parameters between studies in review articles [14, 15, 16, 19, 20, 21, 22, 23, 25], meta-analyses [24, 36, 37], and discussion sections of empirical papers: When model variables are compared between different types of studies, there is an implicit assumption of generalization. Relying on interpretability, model variables have been expected to be associated with specific neural substrates (e.g., reward prediction errors and dopamine function [38]), to expose the core of what differentiates participants with psychiatric conditions from healthy ones (e.g., working-memory parameter differences in schizophrenia [39]), and generally, to capture processes that are particularly "theoretically meaningful" [14].

However, inconsistencies in empirical results are emerging across the developmental [13, 40, 41, 42], clinical [15, 16, 17, 18], cognitive, and neuroscientific literature [24, 36, 37, 43], potentially suggesting a lack of generalizability and/or interpretability, which is also in accordance with different theoretical considerations [27, 28, 44, 45, 46, 47]. Nevertheless, the degree to which parameters generalize between tasks and are interpretable has not been investigated empirically yet (but see [48] for work on parameter reliability).

This was the goal of the current project. We compared the RL parameters fit to the same individuals across different learning tasks in a single study.

We used 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 previously been used to study RL processes [35, 49, 50]. However, like many tasks in the literature, these tasks likely also engaged other cognitive processes, 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 parameters captured in each task (interpretability). We extensively compared and validated the RL models of each task [27, 30, 51], and previously reported the developmental results separately [35, 49, 50].

We found that the RL parameters that reflect decision noise or exploration (inverse decision temperature β, undirected noise ϵ; for model details, see section 4.5) were most consistent within individuals across tasks, suggesting that these parameters were most generalizable. Decision noise/exploration parameters also showed a consistent developmental pattern across subjects, declining from age 8-17. RL learning rate parameters (α_+, α_-), however, were largely inconsistent within individuals across tasks, showing that they did not generalize. Capturing different variance, they likely also reflected different cognitive processes across tasks. Both of these patterns are consistent with patterns that have started to emerge in the existing literature [13]. Behavioral analyses indicated that task differences, and the associated differences in optimal behavior, might underlie these observed parameter discrepancies. These results suggest that past computational findings are not as generalizable as often assumed, and that future research needs to address the reasons of the observed discrepancies to move the field forward.

2. Results

The next section gives a brief overview of the experimental tasks and computational models, before tackling parameter generalizability (section 2.1) and parameter interpretability (section 2.2). Task details are provided in Fig. 1C-E and section 4.4, and computational models and parameter fitting in section 4.5, as well as the original publications [35, 49, 50].

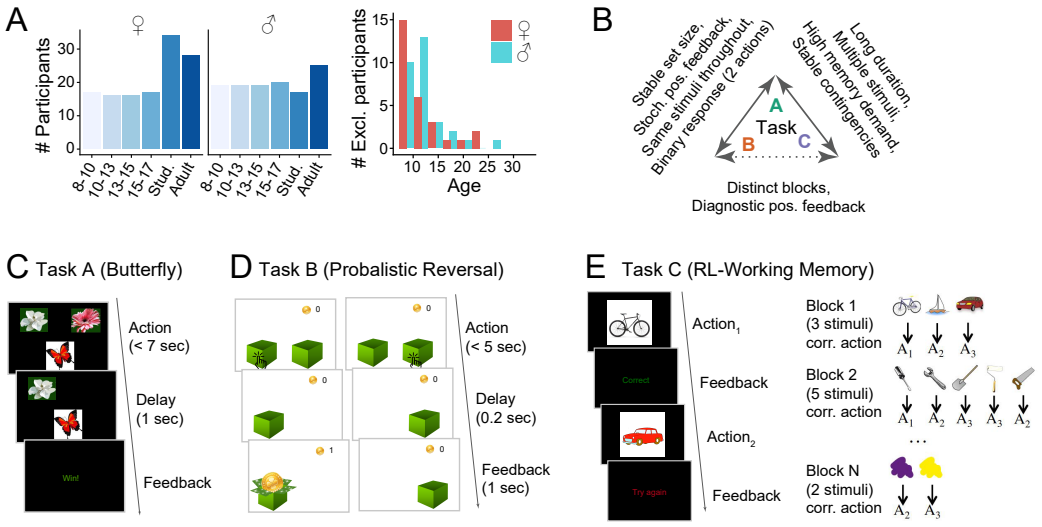

Figure 1: **Overview of the experimental paradigm.** (A) Participant sample. Left: Number of participants in each age group, broken up by sex (self-reported). Age groups were determined by within-sex age quartiles for participants between 8-17 years. The adult sample is broken up by recruitment type ("Stud.": University undergraduates, receiving course credit for participation. "Adult": Adults recruited from the community using the same methods as the developing participants. Right: Number of participants who participated in the study and whose data were excluded because they failed to reach the performance criterion in at least one task. (B) Pairwise similarities in task design between tasks A, B, and C. Similarities between each pair of tasks are shown above the connecting arrows. Only features are shown that differentiate two tasks from the third. E.g., noting "Stable set size" on the edge between tasks A and B implies that set size was not stable in task C. Task A shared more similarities with tasks B and C than they shared with each other. (C) Procedure of task A ("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 [50]. (D) Procedure of task B ("Probababilistic switching"). 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 [49]. (E) Procedure of task C ("Reinforcement learning-working memory"). Participants saw one stimulus on each trial and selected one of three responses. All correct responses and no incorrect responses were rewarded. Stimuli were presented in blocks containing 2-5 different stimuli. The number of stimuli in a block is called 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 [35].

Depending on the task, RL models contained different parameters, reflecting existing differences in the literature. Task A required participants to learn the correct associations between each of four stimuli (butterflies) and two responses (flowers), through 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; and one fixed parameter: learning rate from negative outcomes $\alpha_- = 0$ [50]. 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 [49]. Task C required learning of stimulus-response associations like task A, but over several task blocks with varying numbers of stimuli, and provided deterministic feedback (Fig. 1E). The best model for this task combined RL and working-memory processes, 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 [35, 52].

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, 51]: 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, models' ability to reproduce participants' behavior in simulation, and other criteria of model fit (e.g., interpretability) [44, 45]. We also used hierarchical Bayesian methods for model fitting and comparison when possible to obtain most accurate parameter estimates [51]. Individual publications provide further details [35, 49, 50].

2.1. Part I: Parameter Generalizability

To investigate parameter generalizability, we assessed whether participants showed similar parameter values across tasks, and whether different tasks showed the same parameter age trajectories. These within-participant comparisons are crucial to determine whether discrepancies in the previous literature were caused by methodological differences (e.g., differences in participant samples, testing procedures, modeling quality, research labs), or could arise from mere differences in task characteristics and computational models, as we hypothesized.

2.1.1. Differences in Absolute Parameter Values

We first asked whether tasks led to different absolute parameter values (Fig. 2A), using repeated-measures analyses of variance (ANOVAs). When ANOVAs showed significant task effects, we followed up with pairwise, repeated-measures t-tests, using the Bonferroni correction.

Learning rates α_+ and α_- occupied largely distinct ranges across tasks: Values 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), and fairly high in task B (α_+ mean: 0.77, sd: 0.11; α_- mean: 0.62, sd: 0.14; for statistical comparisons, 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).

All ANOVAs revealed significant and large task effects, and all follow-up t-tests revealed significant and large pairwise differences (Table 1), showing that 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 differences in 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 demands on memory. Using regression models that controlled for age (instead of ANOVA) led to similar results (Table D.9).

2.1.2. Relative Parameter Differences

However, comparing absolute parameter values between tasks has shortcomings: It ignores variance between participants, even though between-participant variance might be the more meaningful measure because it reflects participants' relationships to each other. The simplest way to investigate whether between-participant variance generalized between tasks is to test if individual variance in one task mirrors individual variance in another, 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 task A and tasks B and C, suggesting

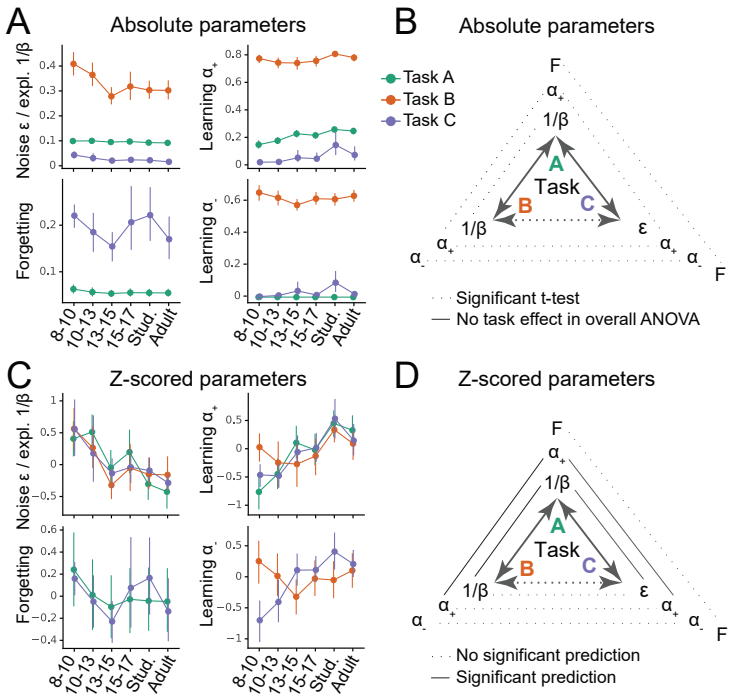

Figure 2: **Generalizability of absolute parameter values (A-B) and of parameter age trajectories / z-scored parameters (C-D) between tasks.** (A) Fitted parameters over participant age (quartile bins), for all three tasks (A: green; B: orange; C: blue). Parameter values differed significantly between tasks (for statistics, see Table 1). 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), visualizing Table 1. Double-sided arrows are replicated from Fig. 1B and show task similarity. Lines show test statistics for absolute parameter values. Dotted lines indicate significant task differences in Bonferroni-corrected pairwise t-tests, which were conducted after observing significant task effects in corresponding ANOVAs. All t-tests were significant, indicating that absolute parameter values differed significantly for each pair of tasks. (C) Parameter age trajectories, i.e., within-task z-scored parameters over participant age bins. Age trajectories can potentially reveal similarities that are obscured by differences in means or variances when assessing absolute parameter values. (D) Summary of the main results of part (C), visualizing Table 4. When parameters in two tasks are connected with a full line, the parameter can be predicted significantly in one task from the other. When parameters are connected with a dotted line, the prediction is not significant. In contrast to absolute parameter values, age trajectories were predictive in several cases, especially for tasks with more similarities (A and B; A and C), compared to tasks with fewer (B and C).

that variance between participants generalized better than absolute values. However, significant correlations were lacking between tasks B and C. This suggests that α_+ and noise/exploration generalized from and to task A, but they did not generalize between tasks B and C, mirroring task similarities (Fig. 1B; also see section 2.2.1; Fig. D.9 shows the correlations between all pairs of features in the dataset.) Note that noise parameters generalized between task 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

However, this correlation analysis is limited in its failure to take into account age effects, a known source of variance, such that apparent task similarities could be driven by a shared dependence on age rather than age-independent underlying similarities. To address this, we next analyzed parameters' age trajectories, which allowed us to abstract away potentially arbitrary differences (e.g., different parameter means and variances across tasks), while conserving potentially meaningful structure in the dataset (i.e., participants' parameter values relative to each other).

We obtained age trajectories by z-scoring each parameter within each task (Fig. 2C). To test for differences in age trajectories, we used mixed-effects regression to predict parameters of all tasks from two age predictors (age and squared age) and task (A, B, or C). When this model fit better than the corresponding model without task, task characteristics affected age trajectories, and we added post-hoc models for each pair of tasks.

For α_-, the task-based regression model showed a significantly better fit, revealing significant task differences (Table 2). Indeed, α_- showed fundamentally different age 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). These differences were reflected in the significant interaction terms of the grand regression model (Table **??**).

For α_+, adding task as a predictor did not improve model fit, suggesting that age trajectories did not differ (Table 2). Indeed, age trajectories were qualitatively similar between tasks, showing linear increases that tapered off with age (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, age trajectories did not differ either (Table 2). For decision noise/exploration, the grand regression model revealed a linear decrease and tapering off in older participants that was consistent across all tasks (Fig. 2C; Table **??**), in accordance with previous findings [13]. For Forgetting, the grand model did not reveal consistent age effects (Fig. 2C; Table **??**).

In summary, when assessing absolute parameter values (Fig. 2A, 2B), differences in scale obscured existing similarities in age trajectories for noise/exploration parameters and α_+ (Fig. 2C). For α_-, on the other hand, differences existed both in terms of scale (Fig. 2A, 2B) and age trajectories (Fig. 2C). As suggested by the correlation analysis, patterns of generalization differed between pairs of tasks, such that more generalization was present between tasks that were more similar in terms of task characteristics (A and B; A and C; not B and C).

2.1.4. Predicting Age Trajectories

So far, we have assessed parameter differences to reveal parameters that do not generalize across tasks. However, the absence of differences only provides indirect evidence *for* generalization. We therefore next assessed how closely parameters were related, using linear regression to predict participants' parameters in one task from the values of the same parameter in a different task. We controlled for age by including age and squared age as predictors to ensure that the prediction was driven by parameter similarities beyond age.

For both α_+ and noise/exploration parameters, task A predicted tasks 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), confirming that α_+ and noise/exploration generalized from and to task A, but not between tasks B and C, mirroring task similarities (Fig. 1B; also see section 2.2.1). For α_-, tasks B and C showed a marginally significant *negative* relationship (Table 4), suggesting that predicting α_- in one task from the other would lead to inverse predictions. Indeed, we previously reported a U-shaped trajectory of α_- in task B with minimum in 13-to-15-year-olds [49], but a consistent increase up to early adulthood in task C [50], revealing striking qualitative differences in the estimation of α_- when using these two tasks. For Forgetting, tasks A

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

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	$p < 0.001$	***

Table 2: Assessing the existence of age effects on parameter trajectories: Model fits 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. The best (significantly smaller) AIC scores are highlighted in bold, and their coefficients are shown in Table **??**.

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	

and C were not predictive of each other (Table 4).

Importantly, these results (Fig. 2D) differ from the previous patterns (Fig. 2C) for Forgetting parameters and α_+ in tasks B and C. This shows that a lack of difference (Fig. 2C) does not imply successful prediction (Fig. 2D).

2.1.5. Summary Part I

In summary, Part I revealed that (1) different tasks led to different estimates of participants' exploration ($\frac{1}{\beta}$), Forgetting (F), and learning rates (α_+, α_-), revealing a lack of generalization of absolute parameter values. Intriguingly, absolute parameter values were stable within tasks (reflecting task demands), but varied within participants. (2) In contrast to absolute parameter values, age trajectories of noise/exploration parameters and learning rates α_+ were qualitatively similar between tasks, suggesting that parameter age trajectories generalized better than absolute values. The age trajectories of learning rates α_-, however, differed fundamentally between tasks, highlighting that parameters in the same models can generalize differently. (3) Assessing the parameters with task-consistent age trajectories,

noise/exploration decreased until early adulthood, in accordance with the literature [13], while learning rates α_+ increased. (4) Using between-task prediction as the strongest test of generalization, age trajectories of learning rates α_- and Forgetting were not predictive, and noise/exploration parameters and learning rates α_+ could only be predicted between similar tasks, suggesting that generalizability was generally weaker than expected, and might depend on task similarity.

2.2. Part II: Parameter Interpretability

Based on these insights, Part II of our investigations focused on parameter interpretability, i.e., the concept that parameters capture specific and unique cognitive processes that are well delineated. We tested parameter interpretability by investigating the relations between different parameters in our dataset, assessing the specificity and distinctiveness of each parameter as well as the relations between parameters and observed patterns of behavior.

2.2.1. The Main Axes of Variation

To gain an understanding of what information was captured by each parameter, we employed a data-driven approach, identifying major axes of variance without specifying a priori hypotheses. We used PCA to identify the major axes in our dataset (composed of both behavioral features and model parameters). We then used these axes (principal components; PCs) to interpret model parameters. To understand the PCs themselves, we analyzed the weights of the behavioral features on each PC (Fig. 3). Detailed information is provided in sections 4.6 (PCA methods), Appendix C (behavioral features), and suppl. Fig. D.10 (additional PCA results).

We first examined PC1, the axis of largest variation (25.1% of explained variance; suppl. Fig. D.10A), to understand the main sources of individual differences in our dataset. Behaviors that indicated good task participation (e.g., high percentage of correct choices) loaded positively on PC1, whereas behaviors that indicated that participants were not on task loaded negatively (e.g., more missed trials, longer response times; Fig. 3A). PC1 comprised measures both in the narrow sense of maximizing task accuracy (e.g., percentage correct choices, measures of task accuracy, win-stay choices), and 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 performance" indicators, reflecting general task engagement. PC1

increased significantly with age, consistent with participants' increasing performance (suppl. Fig. B.6B; age effects of subsequent PCs in suppl. Fig. D.10; suppl. Table D.8).

In all three tasks, noise/exploration loaded negatively on PC1 (Fig. 3A), showing that elevated decision stochasticity was associated with poorer performance in all tasks. Forgetting parameters also loaded negatively, supporting a negative role for performance. α_+ showed positive loadings in all three tasks, suggesting that faster integration of positive feedback was associated with better performance. Intriguingly, 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. This distinction can be interpreted in terms of task demands: Negative feedback was diagnostic in task C, but non-diagnostic in task B (Fig. 1B), such that repeating choices after negative feedback ("Lose-stay" behavior) was hurtful in the former (negative loading on PC1 for task C), but can be beneficial in the latter (positive loading on PC1 for task B; Fig. 3A).

Having gained insight into parameters' roles for task engagement by analyzing PC1, we next turned to PC2 and PC3. To facilitate their interpretation, we flipped the loadings of all PC2 and PC3 features that were negative on PC1, to make them intrepretable with respect to task engagement (for methodological details, see section 4.6). This pre-processing revealed that PC2 and PC3 encoded task contrasts: PC2 contrasted task B to task C (loadings on corresponding features were positive / negative / near-zero for tasks B / C / A; Fig. 3B). PC3 contrasted task A to both B and C (loadings on corresponding features were positive / negative for task A / tasks B and C; Fig. 3C; missed trials and response times did not show task contrasts, suggesting that these features did not differentiate between tasks). The ordering of PC2 and PC3 shows that participants' behavior differed more between tasks B and C (PC2: 8.9% explained variance) than between B or C and A (PC3: 6.2%; suppl. Fig. D.10), in accordance with descriptive task characteristics (Fig. 1B). This shows that after task engagement, the main variation in our dataset arose from task differences.

Intriguingly, noise/exploration parameters, α_+, and α_- reproduced the task contrasts of PC2 and PC3, showing positive or negative loadings based on the task in which they were measured (Fig. 3B, 3C). This means that these parameters differed sufficiently between tasks to be discriminable (as opposed to, e.g., response times and numbers of missed trials, which did not show

task contrasts, suggesting that they were not discriminable between tasks). Each parameter therefore captured enough task-specific variance to make it possible to be identified with the correct task. This degree of differentiability would not be expected if parameters captured the same processes in each task, in which case they would capture the same variance and not show task differences. Taken together, PC2 and PC3 confirmed that each of these parameters captured task-unique processes.

Taken together, the PCA revealed that (1) the main axes of variation in the dataset were task engagement (PC1) and task differences (PC2-PC3). (2) Noise/exploration, Forgetting, α_+, and α_- all were related to task engagement (PC1). Whereas the relation was consistent between tasks for the former three, it was task-dependent for α_- and mirrored specific task demands. (3) Noise/exploration, α_+, and α_- all captured enough task-specific variance to be correctly identified with the corresponding task, showing that they captured different processes depending on the task (PC2-PC3).

2.2.2. Parameters and Cognitive Processes

Whereas the previous analysis revealed that all parameters contained task-specific information, it did not specify how much information was task-specific and how much was shared. For example, noise/exploration parameters contained enough task-specific information to make it possible to determine in which task they were measured (PC2-PC3; Fig. 3B, 3C), but they also showed similar associations with engagement across tasks (PC1; Fig. 3A), similar age trajectories (Fig. 2C), and were mutually predictive (Fig. 2D). To quantify these patterns, we need to understand how much of each parameter's variance was unique and how much was shared between parameters and between tasks.

To achieve this, we probed how much of each parameter's variance was explained by other parameters, using regression. We assumed that parameters reflected one or more cognitive processes, such that shared variance between parameters would imply overlapping cognitive processes. If parameters reflected similar cognitive processes across tasks, then the same parameter should dominate this analysis (e.g., when using parameters in task A to predict $\frac{1}{\beta}$ in task B, task A's $\frac{1}{\beta}$ should show the largest regression coefficient). However, if parameters captured different processes across tasks, this would not be the case (e.g., all parameters of task A might predict task B's $\frac{1}{\beta}$ equally). We used repeated, k-fold cross-validated Ridge regression to avoid overfitting, obtaining unbiased out-of-sample estimates of the

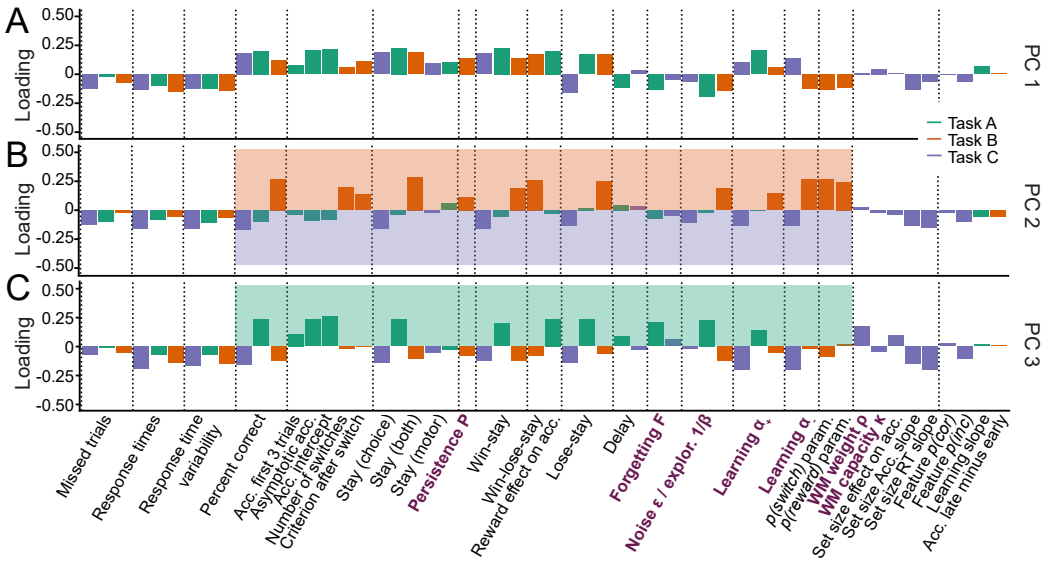

Figure 3: **Identifying the major axes of variation in the dataset.** A PCA was conducted on the entire dataset (39 behavioral features and 15 model parameters). The figure shows the factor loadings of the first three PCs. RL model parameters are highlighted in purple on the x-axis. Behavioral features 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 (blue box). PC3 loaded positively on features of task A (green box) and negatively on features of tasks B and C. We flipped the loadings of features that were negative on PC1 when showing PC2 and PC3 to better visualize the task contrasts (section 4.6).

means and variances of explained variance R^2 and regression coefficients w (for methods, see section 4.7).

We first assessed the overall patterns of prediction, and found that all significant coefficients highlighted shared variance between tasks A and B or tasks A and C, but never between tasks B and C, mirroring our previous results (Fig. 2D; section 2.1.2) and patterns of task similarity (Fig. 1B). This means that no parameters in tasks B or C played a significant role in predicting parameters in the other, while both tasks' parameters were predictive (and being predicted by) parameters in task A. This further highlights the potential role of task similarity in parameter generalizability.

We next focused on noise/exploration parameters. Noise/exploration parameters in tasks B and C showed significant coefficients when predicting noise/exploration in task A, but the inverse was not true, such that noise/exploration in task A did not show significant coefficients when predicting noise/exploration in tasks B or C (Fig. 4A; Table 5). The first result shows that noise/exploration parameters captured variance (cognitive processes) in task A that they also captured in tasks B and C. The second result shows that noise/exploration parameters captured additional cognitive processes in tasks B and C that they did not capture in task A. Furthermore, prediction accuracy increased when combining tasks B and C's parameters to predict noise/exploration in task A, showing that noise/exploration parameters in tasks B and C captured partly non-overlapping aspects of noise/exploration in task A (Fig. 4B, left-most set of bars, compare purple to orange and blue). This highlights both specificity in terms of which cognitive processes were captured by noise/exploration parameters across tasks (prediction between similar tasks), and some lack thereof (prediction was just one-way; no prediction between dissimilar tasks). Furthermore, noise/exploration in task A was predicted 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 the other tasks, revealing a lack of distinctiveness in noise/exploration 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). 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 shows a lack of specificity, implying that learning rates did not reflect a consistent core of cognitive processes across tasks.

We then turned to the distinctiveness of learning rate parameters. 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 sets of parameters in task C [52]. In support of this interpretation, no variance in task C's working-memory parameters could be explained by any other parameters (Fig. 4B), revealing that they captured unique cognitive processes, likely working memory. 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.

2.2.3. Parameters and Behavior

Faced with mounting evidence for parameter inconsistencies, we lastly aimed to uncover whether parameters shared any consistent similarities across tasks. The previous sections showed that parameters likely captured different (neuro)cognitive processes across tasks (e.g., different internal characteristics of learning and choice). However, computational models are fundamentally models of behavior, so we argued that parameters might capture similar behavioral features (e.g., similar tendencies to stay after positive feedback). Even though related, (neuro)cognitive processes and behavioral patterns should not be equated (Fig. 5). For example, different (neuro)cognitive mechanisms (e.g., prefrontal cortical reasoning, basal ganglia value learning, hippocampal episodic memory) might underlie the same behavioral pattern (e.g., lose-stay behavior) in different tasks, depending on the characteristics (e.g., stable versus volatile contingencies; deterministic versus stochastic feedback).

To investigate this possibility, we assessed the relationships between model

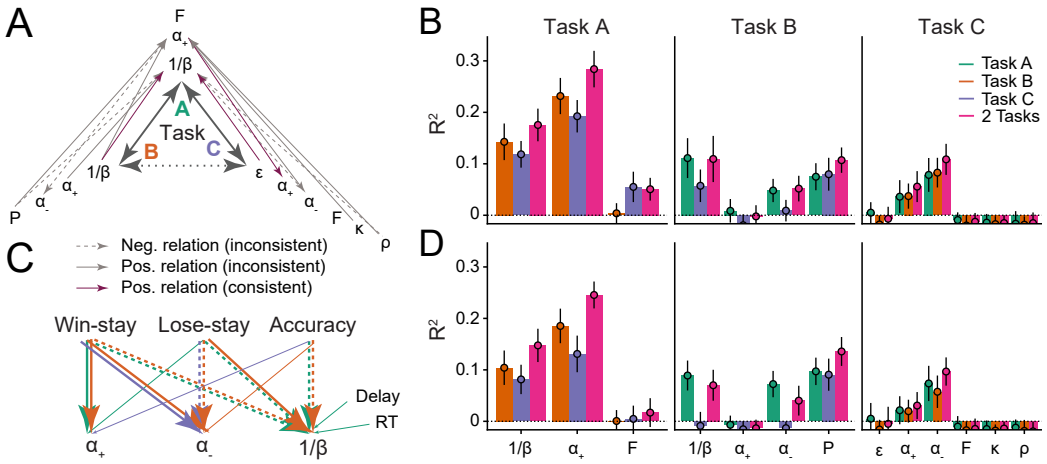

Figure 4: **Assessing parameter interpretability by analyzing shared variance.** (A) Parameter variance (cognitive processing) that is shared between tasks. Each arrow shows a significant regression coefficient when predicting a parameter in one task (e.g., α_+ in task A) from all parameters of a different 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 end. 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) Variance of each parameter that was also 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., multiple arrows). (D) Variance of each parameter that was explained by behavioral features; corresponds to the behavioral Ridge models shown in part (C).

parameters and behavioral features across tasks. Using regularized Ridge regression like above, we predicted each model parameter from five selected behavioral features (Appendix A, Appendix C) of each of the three tasks (15 predictors; for regression methods, see section 4.7). One possible outcome of this analysis is "absolute consistency": parameters might capture the same behavioral pattern within and across tasks (e.g., noise/exploration of each task might capture task A accuracy). This outcome would be expected if parameters captured the same cognitive processes across tasks, and behavioral features were a direct reflection of cognitive processes. Another possible outcome is "absolute *in*consistency" (e.g., in every task, noise/exploration might capture different behavioral features). This outcome would suggest that parameters captured unrelated cognitive and behavioral features in each task. Crucially, a third possible outcome is "task-based consistency": Parameters might capture the same behavioral features, but only within tasks (e.g., in each task, learning rates might capture the win-stay behavior of that task, but not of other tasks). This outcome would suggest that parameters generalized in terms of which behavioral features they reflected, but behavioral features—like (neuro)cognitive processes—differed between tasks.

Focusing on noise/exploration parameters, $\frac{1}{\beta}$ in tasks A and B was predicted by task A win-stay behavior, revealing absolute consistency (Table 6). $\frac{1}{\beta}$ was also predicted by accuracy, win-stay, and lose-stay behavior within both tasks A and B, but not across tasks, revealing task-based consistency (Fig. 4C; Table 6). For learning rates, α_+ in tasks A and B was predicted by the corresponding win-stay behavior, and α_- in tasks B and C was negatively predicted by the corresponding lose-stay behavior, and positively by the corresponding win-stay behavior (Fig. 4C; Table 6), revealing task-based consistency. The consistency of α_- is especially noteworthy given the abundance of discrepancies in previous sections.

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 information (Fig. 3B, 3C), age trajectories (Fig. 2C, 2D), and absolute values (Fig. 2A, 2B). Notably, the observed discrepancies reflected task characteristics (Fig. 1B, 3A) for both parameters (Fig. 1B, 3A) and behavior (suppl. Fig. B.6B), suggesting that task characteristics shaped behavioral responses and model parameters.

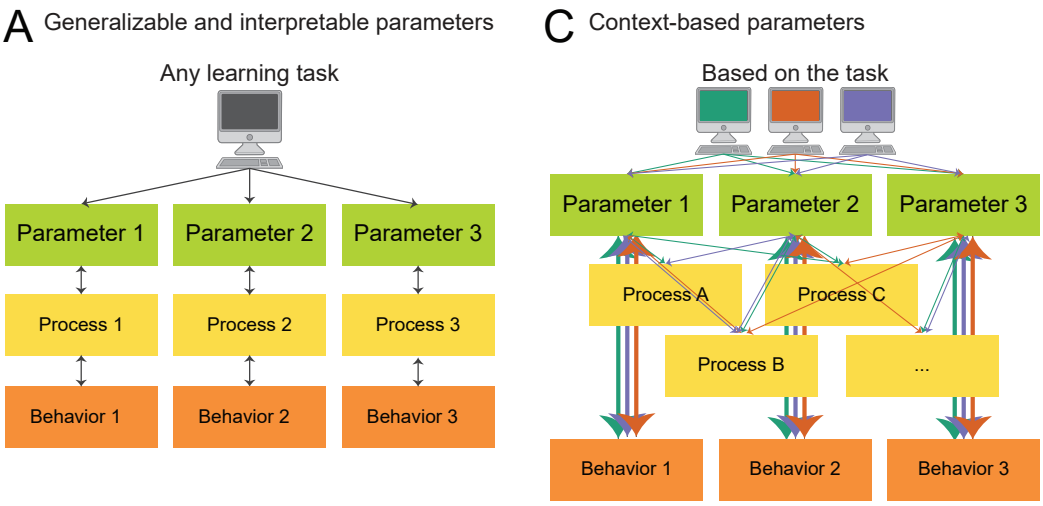

Figure 5: **What do model parameters measure?** (A) View based on generalizability and interpretability. In this view, which is implicitly taken by much current computational modeling research, models are fitted in order to reveal individuals' intrinsic model parameters, which reflect clearly delineated, 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).

3. Discussion

The current study subjected the generalizability and interpretability of RL models to a scrutinous empirical investigation, using a developmental sample. We found weaker levels of generalizability and interpretability than would be expected based on current research practices, such as comparing parameters between studies that use different tasks and models [28].

Interestingly, patterns of generalizability and interpretability varied between parameters: Exploration/noise parameters showed considerable generalizability in the form of correlated variance and age trajectories. The decline in exploration/noise we observed between ages 8-17 was also consistent with previous studies reviewed in [13]. Interpretability of exploration/noise parameters 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, on the other hand, 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 interpretable to a limited degree (overlap in variance between some tasks). However, interpretability was overshadowed by a lack of specificity (lack of shared core variance) and distinctiveness (fundamental entangling with several other parameters, most notably working-memory parameters).

These within-participant findings are consistent with 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 by confirming 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, 53, 54]. 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.

Our findings also help us clarify the source of parameter inconsistencies

in the previous literature, which could indicate replication problems and technical issues: For example, model misspecification [13], lack of model comparison and validation [27, 30], inappropriate fitting methods [29, 51], and lack of parameter reliability due to suboptimal methods [48] have all been suggested as potential sources of inconsistencies. However, our results show that discrepancies are expected even with a consistent methodological pipeline, and using up-to-date modeling techniques (detailed model comparison, validation, and hierarchical Bayesian model fitting where possible). This should encourage the field of computational modeling to study the external factors that drive such inconsistencies, and are currently undescribed by RL methods, with more rigor.

3.1. Limitations

One limitation of our results is that regression analyses might be contaminated by parameter cross-correlations (in sections 2.1.2, 2.1.3, 2.1.4), which would reflect modeling limitations (fewer true degrees of freedom than model 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 correlations between parameters within tasks 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 necessarily and inherently specific, distinct, and "theoretically meaningful" [14] (interpretable). Parameters with the same names also do not automatically transfer between

tasks or models, and are less interchangeable than we often implicitly assume [28]. At the same time, the behavioral features that are captured by parameters seem to 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. **Optimality**. Variance in RL parameters may reflect how participants adapt their behavior to task demands, an explanation proposed by [13]. For example, participants might tune learning rates to task characteristics (e.g., adopting lower learning rates in stable than volatile contexts [55]), rather than learning rates reflecting intrinsic "settings" (e.g., 10-year-olds having a learning rate of 20%; 16-year-olds of 40%). An optimality-based view would also explain why learning rates differ between deterministic and stochastic tasks, which require different amounts of behavioral change in response to feedback, to reach optimal performance. Age differences can potentially be explained because optimal settings likely differ between ages because they interact with different environments, or because different ages might have different capacities to shift internal settings when shifting from task to task. More research is needed, however, to determine whether parameter optimality can explain all inconsistencies in the literature. For example, our finding that participants showed the most optimal parameter values in the intermediate age range in task B [49], whereas optimality increased monotonously with age in tasks A and C [35, 50], is difficult to reconcile with this view.
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) [56, 57, 58]. Some modulatory processes reflect the integration of external contextual information: for example, uncertainty affects dopamine neuron firing [59, 60, 61]. In addition, environments with different degrees of uncertainty have been shown to elicit different learning rates [55]. It is thus possible that neuromodulation by task uncertainty could modulate RL processes, reflected in RL parameters. In our data, feedback stochasticity and task volatility likely contribute to such uncertainty-related modulation. However, other factors like task similarity (low versus high), task characteristics (e.g., volatility

[55, 49], feedback stochasticity, memory load [35, 52], feedback valence and conditioning type [24]), and choice of model parameters (e.g., forgetting [35, 50], counter-factual learning [49], negative and positive learning rates [34, 62, 63]), also seem to affect RL parameters, but are independent of uncertainty. More research is needed to systematically investigate the factors the contribute to modulatory processes, and how they impact cognition and computation.

3. **RL processes are multifaceted.** RL Parameters capture a multitude of separate processes, whose composition 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]. As behavioral models, their use has spanned behaviors from simple conditioning [1, 38] to complex decision making [4, 6, 7, 8, 9, 22, 64], meaning that the same parameters capture cognitive processes that vary considerably in type and complexity: Processes can include the slow acquisition of implicit preferences [1], long-term memory for such preferences [65], quick recognition of contingency switches [49, 66], selection of abstract high-level strategies [7, 9, 67], meta-learning [6], habitual and goal-directed decision making [5], working memory or episodic memory-guided choice [52, 68, 69], and many others. This list alone outnumbers the list of typical RL model parameters, suggesting that RL parameters capture different (combinations of) cognitive processes depending on the paradigm. Similar arguments have also been made for behavioral analyses [70].

3.3. Conclusion

Our research has important implications for fields that focus on individual differences, including developmental and clinical computational research. The current study should be seen as a proof of concept that many contextual factors impact computational modeling, and larger studies will be necessary to quantify these effects and determine their structure. Other areas of modeling besides the RL framework should be subjected to a similar investigation. It is possible, for example, that generalizability differs for sequential sampling [71, 72], Bayesian inference [49, 73, 74], model-based versus model-free RL [48, 75, 76], or other models.

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 [77], 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 [78] may also modulate which cognitive processes we recruit. Heightened awareness and systematic study of contextual variables will therefore be a valuable future investment as we work to measure and accommodate diversity in cognitive processes.

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" [50, 79]), B ("Probabilistic Switching" [66, 49]), and C ("Reinforcement learning-Working memory" [35, 52])—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, 36, 37]. 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 [50]; task B [49]; task C [35]). 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 [80], 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.

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 [80]). 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 [52, 35]. After the second task, participants between the ages of 8-17 provided a saliva sample (for details, see [35]) and took a snack break (5-10 minutes). After that, participants completed task A ("Butterfly task"), which took about 15 minutes [79, 50], and task B ("Probabilistic Switching"), which took about 10 minutes to complete [49]. At the conclusion of the tasks, participants between 11 and 18 completed the Pubertal Development Scale (PDS [81]) 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 [35, 49, 50]. For methodological details, refer to [35]. 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 [50].

4.4.2. Task B ("Probabilistic Switching")

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 [49].

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 [35].

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 [63]. 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 [34, 40, 82, 83, 84, 85, 86, 87, 88]. 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 [89, 90, 91]. 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 [35, 49, 50].

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: [50]; task B: [49]; task C: [35]).

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 [50]. 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" [92] 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 parameters 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.

997 doi:10.1016/j.conb.2012.05.008.
998 URL http://linkinghub.elsevier.com/retrieve/pii/S0959438812000876

999 [9] A. G. E. Collins, E. Koechlin, Reasoning, Learning, and Creativity:
1000 Frontal Lobe Function and Human Decision-Making, PLOS Biology
1001 10 (3) (2012) e1001293. doi:10.1371/journal.pbio.1001293.

1002 [10] D. M. Werchan, A. G. E. Collins, M. J. Frank, D. Amso, Role of
1003 Prefrontal Cortex in Learning and Generalizing Hierarchical Rules
1004 in 8-Month-Old Infants, The Journal of Neuroscience 36 (40) (2016)
1005 10314–10322. doi:10.1523/JNEUROSCI.1351-16.2016.
1006 URL http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1351-16.2016

1007 [11] W. van den Bos, R. Bruckner, M. R. Nassar, R. Mata, B. Eppinger, Computational neuroscience across the lifespan: Promises
1008 and pitfalls, Developmental Cognitive Neuroscience (Oct. 2017).
1009 doi:10.1016/j.dcn.2017.09.008.
1010 URL http://linkinghub.elsevier.com/retrieve/pii/S1878929317301068

1011 [12] F. Bolenz, A. M. F. Reiter, B. Eppinger, Developmental Changes
1012 in Learning: Computational Mechanisms and Social Influences, Frontiers in Psychology 8, publisher: Frontiers (2017).
1013 doi:10.3389/fpsyg.2017.02048.
1014 URL https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02048/full

1015 [13] K. Nussenbaum, C. A. Hartley, Reinforcement learning across
1016 development: What insights can we draw from a decade of research?, Developmental Cognitive Neuroscience 40 (2019) 100733.
1017 doi:10.1016/j.dcn.2019.100733.
1018 URL http://www.sciencedirect.com/science/article/pii/S1878929319303202

1019 [14] Q. J. M. Huys, T. V. Maia, M. J. Frank, Computational psychiatry as a
1020 bridge from neuroscience to clinical applications, Nature neuroscience
1021 19 (3) (2016) 404–413. doi:10.1038/nn.4238.
1022 URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5443409/

1023 [15] R. A. Adams, Q. J. M. Huys, J. P. Roiser, Computational Psychiatry:
1024 towards a mathematically informed understanding of mental illness,
1025 Journal of Neurology, Neurosurgery & Psychiatry 87 (1) (2016) 53–
1026 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. E. Collins, What do Reinforcement Learning Models Measure? Interpreting Model Parameters in Cognition and Neuroscience, psyArxivType: article (May 2021). doi:10.31234/osf.io/e7kwx.
URL https://psyarxiv.com/e7kwx/

[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. 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

[34] 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/

[35] 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

[36] 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

[37] 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/

[38] 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

[39] 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

[40] 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. doi:10.1162/jocn$_{a0}$0677.

[41] 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

[42] 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

[43] 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

[44] 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

[45] 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/article/

[46] 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

[47] 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/

[48] 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

[49] 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

[50] L. Xia, S. Master, M. Eckstein, L. Wilbrecht, A. G. E. Collins, Learning under uncertainty changes during adolescence, in: Proceedings of the Cognitive Science Society, 2020.

[51] 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

[52] 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.

[53] 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.

[54] 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

[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] 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

[57] 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

[58] 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.

[59] 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.Starkweather, S.$

[60] 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

[62] 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

[63] 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

[64] C. Diuk, A. Schapiro, N. Córdova, J. Ribas-Fernandes, Y. Niv, M. Botvinick, Divide and Conquer: Hierarchical Reinforcement Learning and Task Decomposition in Humans, in: Computational and Robotic Models of the Hierarchical Organization of Behavior, Springer, Berlin, Heidelberg, 2013, pp. 271–291. doi:10.1007/978-3-642-39875-9$_1$2.

[65] 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.

[66] 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.

[67] 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

[68] 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

[69] 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

[70] 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.

[71] 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.

[72] 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

[73] 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

[74] 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

[75] 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

[76] 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

[77] 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

[78] 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

[79] 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

[80] 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

[81] 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

[82] 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. doi:10.1162/jocn$_{a0}$0447.

[83] 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

[84] 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.

[85] 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.

[86] 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/

[87] 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

[88] 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

[89] 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.

[90] 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.

[91] 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

[92] 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

[93] 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

[94] 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

[95] 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

[96] 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

[97] 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

[98] 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

[99] 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

[100] 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

[101] 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

[102] 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.

1368 doi:10.1016/j.biopsych.2017.05.017.
1369 URL http://www.sciencedirect.com/science/article/pii/S0006322317316190