Developmental Cognitive Neuroscience 55 (2022) 101106

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

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

Reinforcement learning and Bayesian inference provide complementary models for the unique advantage of adolescents in stochastic reversal

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

[a] Department of Psychology, 2121 Berkeley Way West, USA
[b] Institute of Human Development, 2121 Berkeley Way West, USA
[c] Helen Wills Neuroscience Institute, 175 Li Ka Shing Center, Berkeley, CA 94720, USA

ARTICLE INFO

Keywords:
Reinforcement learning
Bayesian inference
Computational modeling
Development
Volatility
Adolescence
Non-linear changes

ABSTRACT

During adolescence, youth venture out, explore the wider world, and are challenged to learn how to navigate novel and uncertain environments. We investigated how performance changes across adolescent development in a stochastic, volatile reversal-learning task that uniquely taxes the balance of persistence and flexibility. In a sample of 291 participants aged 8–30, we found that in the mid-teen years, adolescents outperformed both younger and older participants. We developed two independent cognitive models, based on Reinforcement learning (RL) and Bayesian inference (BI). The RL parameter for learning from negative outcomes and the BI parameters specifying participants' mental models were closest to optimal in mid-teen adolescents, suggesting a central role in adolescent cognitive processing. By contrast, persistence and noise parameters improved monotonically with age. We distilled the insights of RL and BI using principal component analysis and found that three shared components interacted to form the adolescent performance peak: adult-like behavioral quality, child-like time scales, and developmentally-unique processing of positive feedback. This research highlights adolescence as a neurodevelopmental window that can create performance advantages in volatile and uncertain environments. It also shows how detailed insights can be gleaned by using cognitive models in new ways.

1. Introduction

In mammals and other species with parental care, there is typically an adolescent stage of development in which the young are no longer supported by parental care, but are not yet adult (Natterson-Horowitz and Bowers, 2019). This adolescent period is increasingly viewed as a critical epoch in which organisms explore the world, make pivotal decisions with short- and long-term impact on survival (Frankenhuis and Walasek, 2020), and learn about important features of their environment (Steinberg, 2005; DePasque and Galván, 2017), likely taking advantage of a second window of brain plasticity (Piekarski et al., 2017b; Larsen and Luna, 2018; Lourenco and Casey, 2013).

In humans, adolescence often involves an expansion of environmental contexts (e.g., new pastime activities, growing relevance of peer relationships) and increasingly frequent transitions between such contexts, creating *contextual volatility* (Albert et al., 2013; Somerville et al., 2017). Adolescents also experience increased *outcome stochasticity*, for example as a consequence of increased risk-taking and sensation seeking (Romer and Hennessy, 2007; van den Bos and Hertwig, 2017). Accordingly, it has been argued that adolescent brains and minds may be specifically adapted to contextual volatility and outcome stochasticity, showing an increased ability to learn from and succeed in these situations, compared to both younger and older people (Dahl et al., 2018; Sercombe, 2014; Davidow et al., 2016; Johnson and Wilbrecht, 2011; Lourenco and Casey, 2013; Lloyd et al., 2020).

This prediction would reveal itself as an (inverse) *U-shaped* relationship of age and performance in volatile and stochastic contexts. Indeed, recent research has revealed U-shaped developmental patterns in several related domains, including creativity (Kleibeuker et al., 2013), social learning (Gopnik et al., 2017; Brandner et al., 2021; Blakemore and Robbins, 2012), and value learning (Insel and Somerville, 2018; Rosenbaum et al., 2020; Davidow et al., 2016; Cauffman et al., 2010). However, other studies have shown linear trajectories (i.e., continuous increase from childhood to adulthood, with intermediate levels during adolescence; e.g., (Decker et al., 2016; Xia et al., 2021)) or saturating patterns (i.e., adolescence as crucial period in which adult levels are reached; e.g., (Master et al., 2020; Defoe et al., 2015; Somerville et al., 2017)). It is therefore an open question whether the development of

[*] Corresponding author.
E-mail address: maria.eckstein@berkeley.edu (M.K. Eckstein).

https://doi.org/10.1016/j.dcn.2022.101106
Received 3 September 2021; Received in revised form 1 March 2022; Accepted 25 March 2022
Available online 22 April 2022

decision making follows a linear, saturating, or U-shaped trajectory in volatile and stochastic contexts.

The goal of this study was to examine this question in a controlled laboratory environment. We used a large cross-sectional developmental sample (n = 291) with a wide, continuous age range (8–30 years), offering enough statistical power to observe non-linear effects of age. We also aimed to identify a computational explanation of developmental changes in behavior, using a novel computational modeling approach.

1.1. *Stochastic reversal learning*

To measure learning in volatile and stochastic contexts, we used a stochastic reversal learning task adapted from a two-arm bandit task developed for mice (Tai et al., 2012). Reversal tasks are a cornerstone of cognitive neuroscience, thought to measure response inhibition and cognitive flexibility more broadly (Izquierdo et al., 2017). In stochastic reversal-learning tasks, participants need to balance two opposing goals: On one hand, they need to rapidly change their strategy whenever they identify a context switch; on the other hand, they need to persist in the face of negative outcomes that occur due to feedback stochasticity, but do not signal context switches. Stochastic reversal tasks therefore are sensitive to the balance between persistence and flexibility, which are cognitive capabilities that might undergo crucial development during adolescence (Dahl et al., 2018).

Reversal tasks have been used abundantly in human developmental populations (e.g., (Harms et al., 2018; Dickstein et al., 2010a; Finger et al., 2008; Dickstein et al., 2010b; Hildebrandt et al., 2018; Adleman et al., 2011; Minto de Sousa et al., 2015; DePasque and Galván, 2017)). However, the shape of the developmental trajectory in these tasks remains unclear, based on the current literature. To our knowledge, only three studies have addressed this question directly: Two used binary group designs (assessing *linear* age changes; Javadi et al., 2014; Hauser et al., 2015), but showed no significant age differences. A third study employed a deterministic reversal task, and tested four age groups across adolescence, which allowed to assess *non-linear* changes (van der Schaaf et al., 2011). In this study, an adolescent peak in reversal performance was observed when participants trained with negative feedback, but not positive feedback (Fig. 3; van der Schaaf et al., 2011). Here, we sought to extend this finding by studying a larger sample, adding stochasticity, and providing insights into the cognitive mechanisms that support adolescents' performance by using computational modeling.

An abundance of studies has mapped the brain regions and neurotransmitters that support reversal learning (e.g., orbitofrontal cortex, medial prefrontal cortex, striatum, basal ganglia; serotonin, dopamine, glutamate; Izquierdo et al., 2017; Clark et al., 2004; Izquierdo and Jentsch, 2012; Frank and Claus, 2006; Hamilton and Brigman, 2015; Kehagia et al., 2010; Yaple and Yu, 2019). Many of these neural substrates undergo critical developmental changes during adolescence and early adulthood, often following non-linear trajectories (Toga et al., 2006; Giedd et al., 1999; Sowell et al., 2003; Gracia-Tabuenca et al., 2021; Casey et al., 2008; Somerville and Casey, 2010; Albert et al., 2013; Lourenco and Casey, 2013; DePasque and Galván, 2017; Piekarski et al., 2017b; Dahl et al., 2018; Larsen and Luna, 2018; Laube et al., 2020a). Importantly, some of these neural changes (specifically dopamine and ventral striatum) may be related to behavioral U-shapes (Harden and Tucker-Drob, 2011; Braams et al., 2015), with a potential role for puberty onset (Braams et al., 2015; Op de Macks et al., 2016; Blakemore et al., 2010), as shown by research on sensation seeking and risk taking. Therefore, the developmental trajectory of the neural systems relevant for reversal learning, as well as their behavioral correlates, are in accordance with a special role of adolescence in the current paradigm.

Studies of learning and cognitive flexibility in developing animals may support a U-shaped prediction as well. Rodent studies have revealed differences in adolescent performance under stochastic and volatile conditions. Adolescent rodents showed more robust Pavlovian responding compared to adults when reinforcement is probabilistic, but not deterministic (Meyer and Bucci, 2016). Adolescent rodents showed greater flexibility in reversal learning compared to adult when task contexts involved four choices but not two choices, a factor contributing to greater uncertainty (Johnson and Wilbrecht, 2011). Adolescent rodents have also shown greater flexibility than adults updating responses to a cue that changed from signaling an inhibitory response to an appetitive response (Simon et al., 2013. There is also literature in which adult rodents show performance advantage over adolescents, see Newman and McGaughy, 2011; Shepard et al., 2017). More broadly, human and non-human animal studies both hint that the flux of multiple systems developing simultaneously produces complex effects on learning and decision making (Master et al., 2020), such that linear changes in two or more systems may generate non-linear patterns in behavior. Computational modeling can potentially shed light on such an interplay between systems.

1.2. *Computational modeling*

1.2.1. *Reinforcement Learning (RL)*

RL is a popular framework to model probabilistic reversal learning (Gläscher et al., 2009; Peterson et al., 2009; Chase et al., 2010; Javadi et al., 2014; Hauser et al., 2015; Boehme et al., 2017; Metha et al., 2020). RL agents choose actions based on action *values* that reflect actions' expected long-term cumulative reward. Action values are typically estimated by incrementally updating them every time an action outcome is observed (Section 4.5.1). The size of each update, determined by an agent's *learning rate*, captures the integration time scale, i.e., whether value estimates are based on few recent outcomes, or many outcomes that reach further into the past. A specialized network of brain regions, including the striatum and frontal cortex, has been associated with specific RL-like computations (Frank and Claus, 2006; Niv, 2009; Lee et al., 2012; O'Doherty et al., 2015).

As a computational model, RL interprets cognitive processing during reversal learning as *value learning*: RL agents continuously adjust current action values based on new outcomes, striving to learn increasingly accurate values (Fig. 3A, left). Importantly, the same gradual learning process occurs during stable task periods and after context switches, without an explicit concept of switching. Behavioral switching only occurs once the previously-rewarding action has accumulated enough negative outcomes to push its value below the previously-unrewarding action. This gradual change may fail to capture the quick and flexible switching behavior observed in humans and non-human animals (Izquierdo et al., 2017; Costa et al., 2015).

Because basic RL algorithms hence behave sub-optimally in volatile environments (Gershman and Uchida, 2019; Sutton and Barto, 2017), we implemented model augmentations that alleviate these issues, including separating learning rates for positive versus negative outcomes (e.g., (van den Bos et al., 2012; Frank et al., 2004; Cazé and van der Meer, 2013; Christakou et al., 2013; Harada, 2020; Palminteri et al., 2016; Javadi et al., 2014; Lefebvre et al., 2017; Dabney et al., 2020)), allowing for counter-factual updating (i.e., learning about non-chosen options; e.g., (Boorman et al., 2011; Palminteri et al., 2016; Boehme et al., 2017; Gläscher et al., 2009; Hauser et al., 2014)), and for choice persistence (i.e., repeating actions independent of the outcome; e.g., (Sugawara and Katahira, 2021)). See Section 4.5.1 for details.

1.2.2. *Bayesian Inference (BI)*

However, many have argued that a different computational framework, BI (specifically, Hidden Markov Models), provides a better model for human and animal behavior in reversal tasks than RL (Gershman and Uchida, 2019; Fuhs and Touretzky, 2007; Bromberg-Martin et al., 2010; Costa et al., 2015; Solway and Botvinick, 2012). Indeed, BI models have shown better fit than RL models in empirical studies on macaques (Bartolo and Averbeck, 2020) and human adults (Hauser

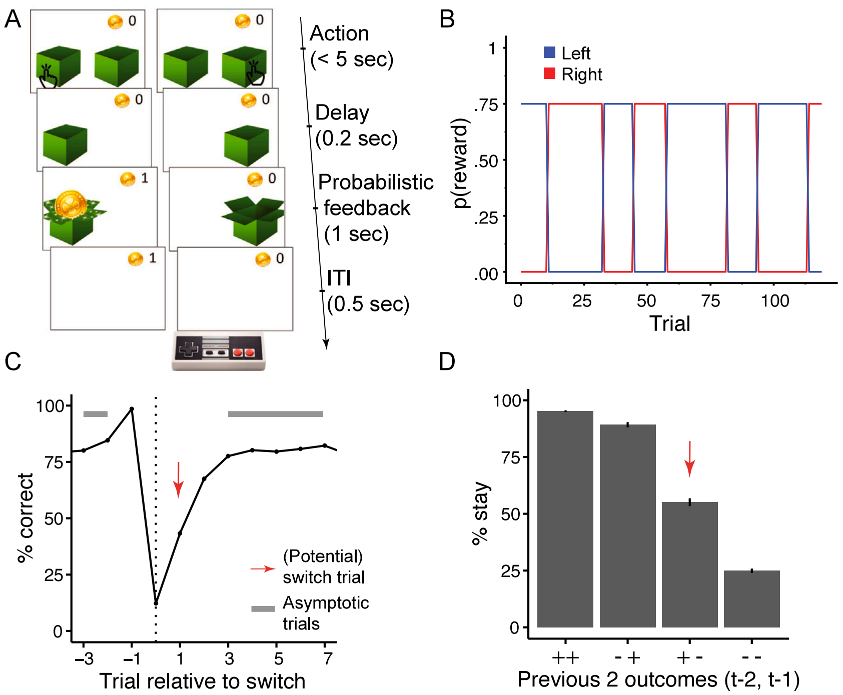

Fig. 1. (A) Task design. On each trial, participants chose one of two boxes, using the two red buttons of the shown game controller. The chosen box either revealed a gold coin (left) or was empty (right). The probability of coin reward was 75% on the rewarded side, and 0% on the non-rewarded side. (B) The rewarded side changed multiple times, according to unpredictable task switches, creating distinct task blocks. Each colored line (blue, red) indicates the reward probability ($p(reward)$) of one box (left, right) at a given trial, for an example session. (C) Average human performance and standard errors, aligned to true task switches (dotted line; trial 0). Switches only occurred after rewarded trials (Section 4.3), resulting in performance of 100% on trial -1. The red arrow shows the switch trial, gray bars show trials included as asymptotic performance. (D) Average probability of repeating a previous choice ("stay") as a function of the two previous outcomes ($t-2$, $t-1$) for this choice ("+": reward; "–": no reward). Error bars indicate between-participant standard errors. Red arrow highlights potential switch trials like in part C, i.e., when a rewarded trial is followed by a non-rewarded one, which—from participants' perspective—is consistent with a task switch. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

et al., 2014; Schlagenhauf et al., 2014; albeit with mixed results in adolescents: Jepma et al., 2020). Furthermore, BI is the standard modeling framework in the "inductive reasoning" literature, whose tasks often have the same structure as stochastic reversal-learning tasks (e.g., (Nassar et al., 2012; O'Reilly et al., 2013; Yu and Dayan, 2005)).

The main reason for the supposed superiority of BI in reversal learning is the ability to reason about *hidden states* and switch behavior rapidly after recognizing state changes. Hidden states are unobservable features that determine an environment's underlying mechanics (e.g., in reversal tasks, which choices are objectively correct and incorrect). These states can be difficult to infer because observable outcomes are often probabilistic. BI agents infer hidden states by engaging *predictive models* that determine how likely different outcomes occur in each state (e.g., how likely a negative outcome occurs after a correct versus incorrect choice). Agents continuously combine state *likelihoods* with their *prior beliefs* about hidden states to obtain updated *posterior beliefs* about these hidden states (Perfors et al., 2011; Sarkka, 2013).

Even though the BI framework therefore provides an excellent choice to model stochastic reversal learning, it is still less common, and hence could provide insights into the development of reversal learning that have so far escaped our attention: For example, BI can characterize predictive mental models and inferential reasoning.

The goal of this study was to characterize adolescent behavior in stochastic reversal, and to identify its underlying cognitive mechanisms: Whereas RL can tell us about participants' learning rates in different situations and is in line with previous developmental modeling work (Javadi et al., 2014; Hauser et al., 2015), as well as the majority of non-developmental work on reversal learning and most standard cognitive neuroscience tasks, BI can assess participants' mental task models and inferential processes, and is increasingly seen as a superior model compared to RL for reversal paradigms.

Using in-depth modeling analyses, we found that both models provided good models of human behavior, and that their insights could be combined to identify features of cognitive processing that went beyond any specific model, including behavioral quality (i.e., task performance in the most general sense) and time scales (i.e., whether decisions are based on just the most recent outcomes, or on a long-running average of many past outcomes). Our results support the existence of an adolescent performance peak in stochastic reversal learning, which can be explained in terms of multiple cognitive mechanisms including learning, exploration, and inference.

2. Results

2.1. Task design

Participants were told that their goal was to collect gold coins, which were hidden in one of two locations (Fig. 1A). The location that contained the coin changed unpredictably, generating *volatility*, and the correct location did not always provide coins, adding *stochasticity*. On each trial, two identical boxes appeared on the screen. Participants chose one, either receiving a coin (reward) or not (Fig. 1A). The correct location was rewarded in 75% of the trials on which it was chosen, whereas the other one was never rewarded, in accordance with the rodent task from which the current task was adapted (Tai et al., 2012). Positive outcomes were therefore diagnostic of correct actions, whereas negative outcomes were ambiguous, arising from either stochastic noise or task switches. After reaching a non-deterministic performance criterion (Section 4.3), an unsignaled switch occurred, and the opposite location became rewarding. Participants encountered 5–9 switches and completed a total of 120 trials (Fig. 1B). Before the main task, participants completed a child-friendly tutorial (Section 4.3).

2.2. Task behavior

Participants gradually adjusted their behavior after task switches, and on average started selecting the correct action about 2 trials after a switch, reaching an asymptotic performance of around 80% correct choices within 3–4 trials after a switch (Fig. 1C). Participants almost always repeated their choice ("stayed") after receiving positive outcomes ("− +" and "+ +"), and often switched actions after receiving two negative outcomes ("− −"). Behavior was most ambivalent after receiving a positive followed by a negative outcome ("+ −"), i.e., on "potential" switch trials (red arrows in Fig. 1C and D; for age differences, see suppl. Fig. 4).

2.2.1. Age differences: Performance peak in adolescents

Several behavioral measures indicate good performance on this task: Overall accuracy (percentage of trials on which the currently experimenter-defined correct box is chosen, independent of reward); total number of points won; response times on correct trials (efficiency of execution); number of blocks completed (because the switch criterion was performance dependent); willingness to repeat a choice ("stay") after a potential switch (signaling an understanding of reward stochasticity); and asymptotic accuracy (accuracy during the "stable", non-switching phase of each block). Some of these measures are correlated and potentially redundant (e.g. accuracy, points, and number of blocks); some are correlated but target different aspects of performance (e.g. win-stay and lose-switch); others can be considered more independent (e.g. reaction times).

We used (logistic) mixed-effects regression to test the continuous effects of age on each performance measure (Section 4.4), and found positive linear and negative quadratic age effects in all cases (Table 1). This is in accordance with a general increase in performance from childhood to adulthood that is modified by the hypothesized adolescent peak in performance. To confirm the existence of this peak (as opposed to any other non-linear developmental pattern that would be reflected in a quadratic age effect), we conducted two-line regression, a method specifically designed to detect U-shapes (Simonsohn, 2018). Indeed, we observed statistically significant U-shapes for overall accuracy, number of points won, response times on correct trials, and number of blocks (Table 2). Except for response times, all of these measures showed change points between 13.29 and 14.55 years of age. Even though the U-shape failed to reach significance for staying after (pot.) switch and asymptotic performance, these measures were also qualitatively consistent with a U-shape and switch point in mid-adolescence.

To further assess the performance peak, we calculated rolling performance averages (Section 4.4), confirming peaks at around 13–15 years in most performance measures, including overall accuracy (Fig. 2A), points won (Fig. 2A), performance after potential switch trials (Fig. 2E), and asymptotic performance (Fig. 2F). Overall accuracy inclined steeply between ages 8–14, after which it gradually declined, settling into a stable plateau around age 20 (Fig. 2A). The number of points showed a similar pattern (Fig. 2B). The willingness to repeat previous actions after a single negative outcome (Fig. 2E) showed a similarly striking increase between children and adolescents, and a (less pronounced) decline for adults. This measure reveals that in our task, adolescents were most persistent in the face of negative feedback. Performance during stable task periods (accuracy on asymptotic trials) also was highest in adolescents, especially compared to younger participants (Fig. 2F). Response times were the only performance measure in which adolescents were outperformed by adult participants (Figs. 2C, 3D).

For easier visualization of finer aspects of behavior, we finally binned participants into discrete age groups, forming four equal-sized bins for participants aged 8–17, and two for adults (section 6.2; suppl. Fig. 11A). The performance peak was again evident in the intermediate age range (third youth quartile), suggesting that mid-adolescents outperformed younger participants, older teenagers, and adults (Fig. 3C–F). This result was not contingent on the choice of binning (Appendix 6.3.3). Repeated, post-hoc, 5-wise Bonferroni-corrected t-tests revealed several significant differences comparing 13-to-15-year-olds to younger and older participants (Fig. 3C–F, suppl. Table 9).

Table 1
Statistics of mixed-effects regression models predicting performance measures from sex (male, female), age (z-scored; "lin."), and quadratic age (square of z-scored age; "qua."; for details, see Section 4.4). Overall accuracy, stay after potential (pot.) switch, and asymptotic performance were modeled using logistic regression, and z-scores are reported. Log-transformed response times on correct trials and total points won were modeled using linear regression, and t-values are reported. * $p < .05$; ** $p < .01$, *** $p < .001$. All models showed significant quadratic effects of age, supporting an inverse-U shaped developmental trajectory of performance.

Performance measure (Figure)	Predictor	β	z/t	p	sig.
Overall accuracy (2A)	Age (z, lin.)	0.043	2.38	0.017	**
	Age (z, qua.)	−0.052	−3.11	0.0019	**
	Sex	0.009	0.2	0.77	
Total points (2B)	Age (z, lin.)	1.23	2.82	0.0052	**
	Age (z, qua.)	−0.036	−3.11	0.0021	**
	Sex	0.19	0.23	0.82	
Response times (2C)	Age (z, lin.)	−0.21	−10.1	< 0.001	***
	Age (z, qua.)	0.14	7.3	< 0.001	***
	Sex	0.19	5.0	< 0.001	***
Number of blocks (2D)	Age (z, lin.)	0.13	2.6	0.0097	**
	Age (z, qua.)	−0.0036	−2.7	0.0070	**
	Sex	0.04	0.4	0.66	
Stay after (pot.) switch (2E)	Age (z, lin.)	0.44	3.78	< 0.001	***
	Age (z, qua.)	−0.38	−3.48	< 0.001	***
	Sex	0.26	1.24	0.21	
Asymptotic perf. (2F)	Age (z, lin.)	0.17	3.57	< 0.001	***
	Age (z, qua.)	−0.18	−3.97	< 0.001	***
	Sex	0.030	0.35	0.73	

Table 2
Two-line regression for performance measures (Simonsohn, 2018). "Slope 1" and "slope 2" indicate the slopes of the younger and older group, respectively. "Age" indicates the age cut off that separates the younger and older group of participants. P-values and significance are reported separately for each group. A U-shaped relationship is present when the slopes of the younger and older groups are both significant, with opposite signs. The existence of such a U-shape is indicated by a star in the "U" column.

Performance measure	Slope 1	p	sig.	Slope 2	p	sig.	Age	U
Overall accuracy	0.91	0.002	**	−0.16	0.028	*	13.29	*
Points won	1.51	< 0.001	***	−0.3	0.002	**	13.66	*
RTs (correct trials)	−20.53	< 0.001	***	6.63	0.030	*	18.92	*
Number of blocks	0.11	0.002	**	−0.03	0.043	*	14.55	*
Stay after (pot.) switch	5.15	< 0.001	***	−0.35	0.53	–	15.13	–
Asympt. perf.	0.01	0.0007	***	0	0.36	–	14.74	–

We next focused on the differential effects of positive compared to negative outcomes on behavior, finding that adolescents adapted their choices more optimally to previous outcomes than younger or older participants. To show this, we used mixed-effects logistic regression to predict actions on trial t from predictors that encoded positive or negative outcomes on trials $t-i$, for delays $1 \leq i \leq 8$ (Section 4.4). First, we observed that the effects of positive outcomes were several times larger than the effects of negative outcomes (suppl. Table 8; Fig. 8B, C, E, F). This pattern was expected given that positive outcomes were diagnostic, whereas negative outcomes were ambivalent, and shows that participants adjusted their behavior accordingly.

The regression also showed an interaction between age and previous outcomes, revealing that the effects of previous outcomes on future behavior changed with age (suppl. Fig. 8B, C, E, F; suppl. Table 8). On trials t − 1 and t − 2, positive outcomes interacted with age and squared age (all $p's < 0.014$; suppl. Table 8), confirming that the effect of positive outcomes increased with age and then slowly plateaued (suppl. Fig. 8C, F). For negative outcomes, the sign of the interaction was opposite for trials $t-1$ versus $t-2$ (all $p's < 0.046$; suppl. Table 8), showing that the effect of negative outcomes flipped, being weakest in adolescents for trial t − 1 (suppl. Fig. 8B), but strongest for trial t − 2. In other words, mid-adolescents were best at ignoring single,

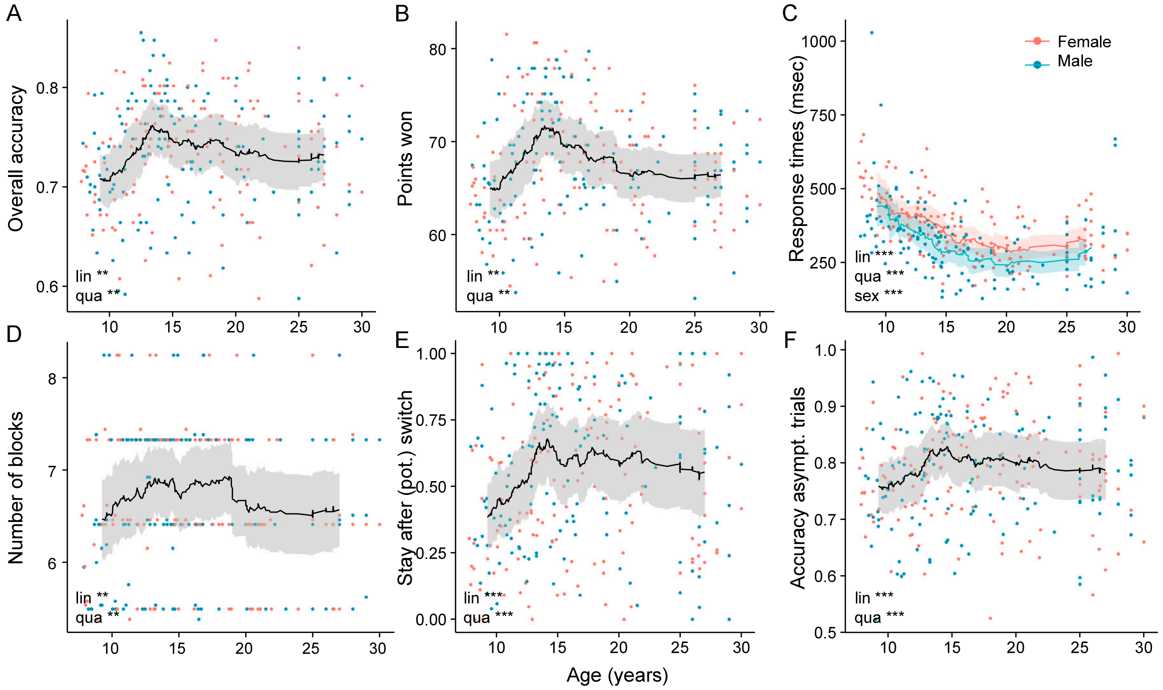

Fig. 2. Task performance across age. Each dot shows one participant, color denotes sex. Lines show rolling averages, shades the standard error of the mean. The stars for "lin", "qua", and "sex" denote the significance of the effects of age, squared age, and sex on each performance measure, based on the regression models in Table 1 (* $p < .05$, ** $p < .01$, *** $p < .001$) (A) Proportion of correct choices across the entire task (120 trials), showing a peak in adolescents. The non-linear development was confirmed by the quadratic effect of age ("qua"; Table 1), and the U-shape by the significant two-line analysis (Table 2). (B) (Corrected) number of points won in the game (Section 4.4), showing a peak in adolescents, confirmed by the quadratic effect of age and significant two-line analysis. (C) Median response times on correct trials. Regression coefficients differed significantly between males and females; rolling averages are hence shown separately. The performance peak occurred after adolescence. (D) (Corrected) number of blocks completed by each participant (Section 4.4), showing a quadratic effect of age. (E) Fraction of trials on which each participant "stayed" after a (potential, "pot.") switch trial (red arrows in Fig. 1C and D), showing a peak in adolescents and quadratic age effect. (F) Accuracy on asymptotic trials (horizontal gray bars in Fig. 1C), also showing a peak in adolescents and quadratic age effect.

ambivalent negative outcomes ($t-1$), but most likely to integrate long-range negative outcomes ($t-2$), which potentially indicate task switches.

To summarize, mid-adolescents outperformed younger participants, older adolescents, and adults on a stochastic reversal task. Performance advantages were evident in several measures of task performance, and likely related to how participants responded to positive and negative outcomes. To understand better which cognitive processes underlay these patterns, we employed computational models featuring RL and BI.

2.3. Cognitive modeling

We first identified a winning model of each family (RL, BI), comparing numerical fits (WAIC; Watanabe, 2013) between basic and augmented implementations within each family (suppl. Fig. 18 and 19; Table 3).

The winning RL model had four free parameters: persistence p, inverse decision temperature β, and learning rates α_+ and α_- for positive and negative outcomes, respectively (Section 4.5.1). In addition to "factual" action value updates on chosen actions, this model also performed "counterfactual" updates on the values of unchosen actions (Palminteri et al., 2016). For example, after receiving a reward for choosing left (factual outcome), the algorithm both decreases the value of the right choice (counterfactual update), and increases the value of the left choice (factual update). The size of both factual and counterfactual updates was controlled by learning rates $\alpha+$ and α_-, simplifying the model (Table 3). Parameters p and β controlled the translation of RL values into choices: Increasing persistence p increased the probability of repeating actions independently of action values. Small β induced decision noise (increasing exploratory choices), and large β allowed for reward-maximizing choices.

The winning BI model also had four parameters: besides choice-parameters p and β like the RL model, these were task volatility p_{switch} and reward stochasticity p_{reward}, which characterized participants' internal task model (Fig. 3A; Section 4.5.2). p_{switch} could represent a stable ($p_{switch} = 0$) or volatile task ($p_{switch} > 0$), and p_{reward} deterministic ($p_{reward} = 1$) or stochastic outcomes ($p_{reward} < 1$). Because the actual task was based on parameters $p_{switch} = 0.05$ and $p_{reward} = 0.75$, an optimal agent would use these values to obtain the most accurate inferences.

In addition to providing better fit (Table 3), the two winning models also validated better behaviorally compared to simpler versions, closely reproducing human behavior (Figs. 4; 3C, E, F; suppl. Fig. 18 and Fig. 19; Palminteri et al., 2017; Wilson and Collins, 2019). The winning RL model had the overall lowest WAIC score, revealing best quantitative fit, but both models validated equally well qualitatively: Both showed behavior that was almost indistinguishable from humans (Fig. 4), and reproduced all qualitative age differences, including adolescents' peak in overall accuracy (Fig. 3C), proportion of staying after (potential) switch trials (Fig. 3E), asymptotic performance on non-switch trials (Fig. 3F), and their most efficient use of previous outcomes to adjust future actions (suppl. Fig. 8B, C, E, F). Other models did not capture all these qualitative patterns (suppl. Fig. 18 and Fig. 19). The closeness in WAIC scores (Table 3) and the equal ability to reproduce details of human behavior reveal that both models captured human behavior adequately, and suggest that both provide plausible explanations of the underlying cognitive processes. We therefore fitted both to participant data to estimate individual parameter values, using hierarchical Bayesian fitting (Fig. 3B; Section 4.5.3).

2.3.1. Age differences in model parameters

Across models, three parameters showed non-monotonic age trajectories, mirroring behavioral differences: α_-, p_{reward}, and p_{switch} declined

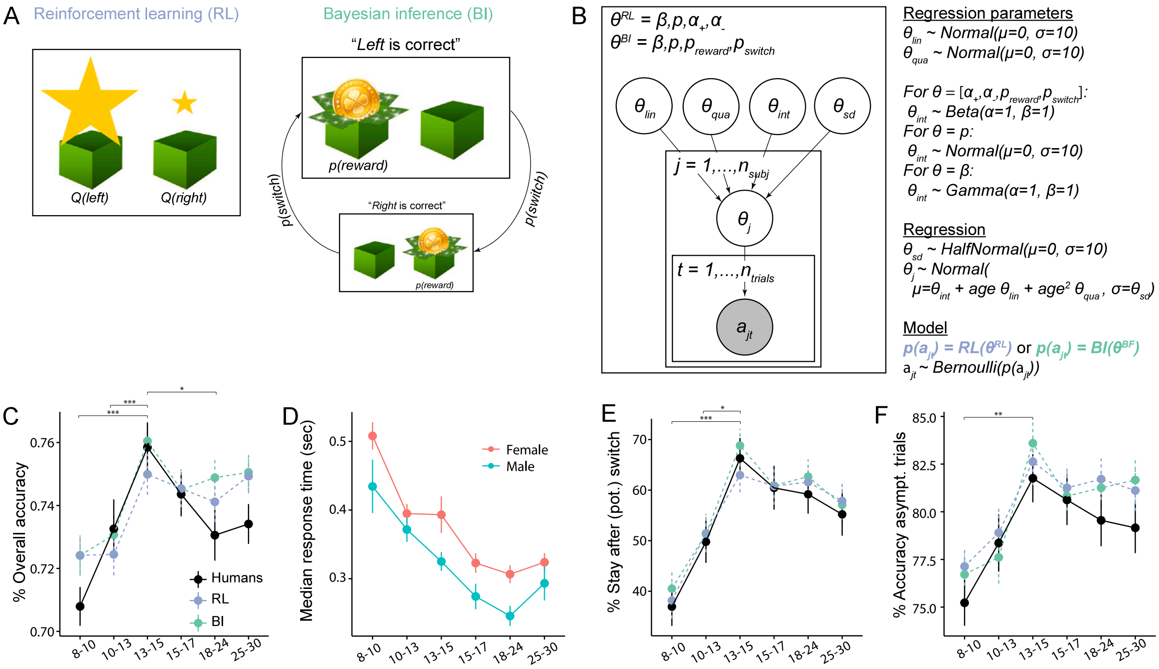

Fig. 3. (A) Conceptual depiction of the RL and BI models. In RL (left), actions are selected based on learned values, illustrated by the size of stars ($Q(left)$, $Q(right)$). Values are calculated based on previous outcomes (Section 4.5.1). In BI (right), actions are selected based on a mental model of the task, which differentiates different hidden states ("Left is correct", "Right is correct"), and specifies the transition probability between them ($p(switch)$) as well as the task's reward stochasticity ($p(reward)$). The sizes of the two boxes illustrate the inferred probability of being in each state (Section 4.5.2). (B) Hierarchical Bayesian model fitting. Box: RL and BI models had free parameters θ^{RL} and θ^{BI}, respectively. Individual parameters θ_j were based on group-level parameters θ_{sd}, θ_{int}, θ_{lin}, and θ_{qua} in a regression setting (see text to the right). For each model, all parameters were simultaneously fit to the observed (shaded) sequence of actions a_{jt} of all participants j and trials t, using MCMC sampling. Right: We chose uninformative priors for group-level parameters; the shape of each prior was based on the parameter's allowed range. For each participant j, each parameter θ was sampled according to a linear regression model, based on group-wide standard deviation θ_{sd}, intercept θ_{int}, linear change with age θ_{lin}, and quadratic change with age θ_{qua}. Each model (RL or BI) provided a choice likelihood $p(a_{jt})$ for each participant j on each trial t, based on individual parameters θ_j. Action selection followed a Bernoulli distribution (for details, see Sections 4.5.3 and 6.2.2). (C)–(F) Human behavior for the measures shown in Fig. 2, binned in age quantiles. (C), (E), and (F) also show simulated model behavior for model validation, verifying that models closely reproduced human behavior and age differences.

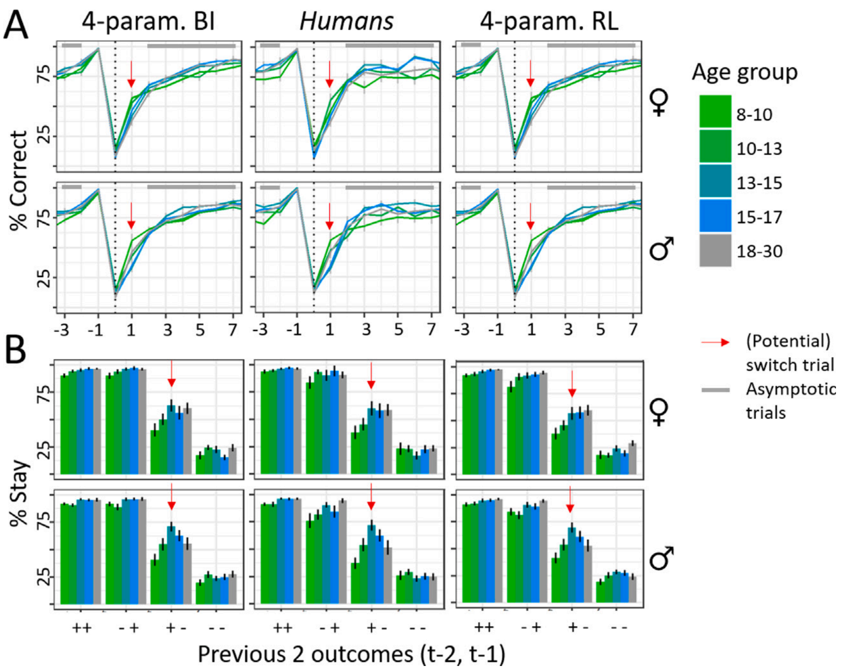

Fig. 4. Model validation, comparing simulated behavior of the winning 4-parameter BI model (left column), humans (middle column), and the winning 4-parameter RL model (right column). Both models show excellent fit, evident in the fact that simulated behavior is barely distinguishable from human behavior (Appendix 6.3.6). (A) Behavior in trials surrounding a real switch of the correct choice ($t = 0$) shows that both models capture well the quick adaptation for all groups. Colors refer to age groups, red arrows show switch trials, gray bars trials of asymptotic performance, like in Fig. 1C. (B) Stay probability in response to outcomes 1 and 2 trials back, like in Fig. 1D. Both models capture well the empirical pattern of switching behavior.

Table 3
WAIC model fits and standard errors for all models, based on hierarchical Bayesian fitting. Bold numbers highlight the winning model of each class. For the parameter-free BI model, the Akaike Information Criterion (AIC) was calculated precisely. WAIC differences are relative to next-best model of the same class, and include estimated standard errors of the difference as an indicator of meaningful difference. In the RL model, "α" refers to the classic RL formulation in which $\alpha_+ = \alpha_-$. "α_c" refers to the model in which factual and counterfactual learning rates were separate, but positive and negative outcomes were not differentiated ($\alpha_{+c} = \alpha_{-c}$; Section 4.5.1).

	Free parameters (count)		(W)AIC	WAIC difference
BI	–	(0)	31,959	2668 ± 0
	β	(1)	$29,291 \pm 206$	868 ± 78
	$\beta,\ p$	(2)	$28,423 \pm 201$	4769 ± 132
	$\beta,\ p,\ p_{reward}$	(3)	$23,654 \pm 203$	51 ± 10
	$\beta,\ p,\ p_{reward},\ p_{switch}$	(4)	**23,603 ± 200**	0
RL	$\alpha,\ \beta$	(2)	$26,678 \pm 200$	438 ± 44
	$\alpha,\ \beta,\ \alpha_c$	(3)	$26,240 \pm 201$	1429 ± 78
	$\alpha,\ \beta,\ \alpha_c,\ p$	(4)	$24,811 \pm 190$	42 ± 13
	$\alpha_+,\ \beta,\ \alpha_{+c},\ p,\ \alpha_-$	(5)	$24,769 \pm 213$	1260 ± 73
	$\alpha_+,\ \beta,\ \alpha_{+c},\ p,\ \alpha_-,\ \alpha_{-c}$	(6)	$23,509 \pm 211$	17 ± 10
	$\alpha_+ = \alpha_{+c},\ \alpha_- = \alpha_{-c},\ \beta,\ p$	(4)	**23,492 ± 201**	0

drastically within the first three age bins (8–13 years), then reversed their trajectory and increased again, reaching slightly lower plateaus around 15 years that lasted through adulthood (Fig. 5). For p_{switch}, age differences were captured in a significant quadratic effect of age in the age-based model (suppl. Table 14; for explanation of age-based and age-less model, see Section 4.5.3). For α_- and p_{reward}, differences were captured in significant pairwise differences between mid-adolescents and other age groups, tested within the age-less model (suppl. Table 13). The two-line regression did not reveal a significant U-shape for these parameters (suppl. Table 15).

BI's mental model parameters p_{switch} and p_{reward} reflect task volatility and stochasticity (Fig. 1A), and can be compared to the true task parameters ($p_{reward} = 0.75$; $p_{switch} = 0.05$) to assess how optimal participants' inferred models were. Both parameters were most optimal in mid-adolescents, whereas younger and older participants strikingly overestimated volatility (larger p_{switch}), while underestimating stochasticity (larger p_{reward}). Similarly in RL, α_- was lowest in mid-adolescents. Indeed, lower learning rates for negative feedback α_- were beneficial because they avoided premature switching based on single negative outcomes, while allowing adaptive switching after multiple negative outcomes.

In both RL and BI, choice parameters p and β increased monotonically with age, growing rapidly at first and plateauing around early adulthood (Fig. 5, top two rows). The age-based model (Section 4.5.3) revealed that both the linear and negative quadratic effects of age were significant (suppl. Table 14). This shows that participants' willingness to repeat previous actions independently of outcomes (p) and to exploit the best known option (β) steadily increased until adulthood, including steady growth during the teen years. Results of the two-line regression confirmed the monotonic increase in these parameters, as well as the later maturation (suppl. Table 15). Parameter α_+ showed a unique stepped age trajectory, featuring relatively stable values throughout childhood and adolescence, and an increase in adults (Fig. 5, left column).

Through the lens of RL, these findings suggest that adolescents outperformed other age groups because they integrated negative feedback more optimally (α_-). Through the lens of BI, the performance peak occurred because adolescents used a more accurate mental task model (p_{switch} and p_{reward}). Taken together, both models agree that behavioral differences arose from cognitive difference in the "update step" of feedback processing (i.e., value updating in RL; state inference in BI). Age differences in the "choice step" (i.e., selecting actions), however, showed monotonous age differences with steady growth in adolescents.

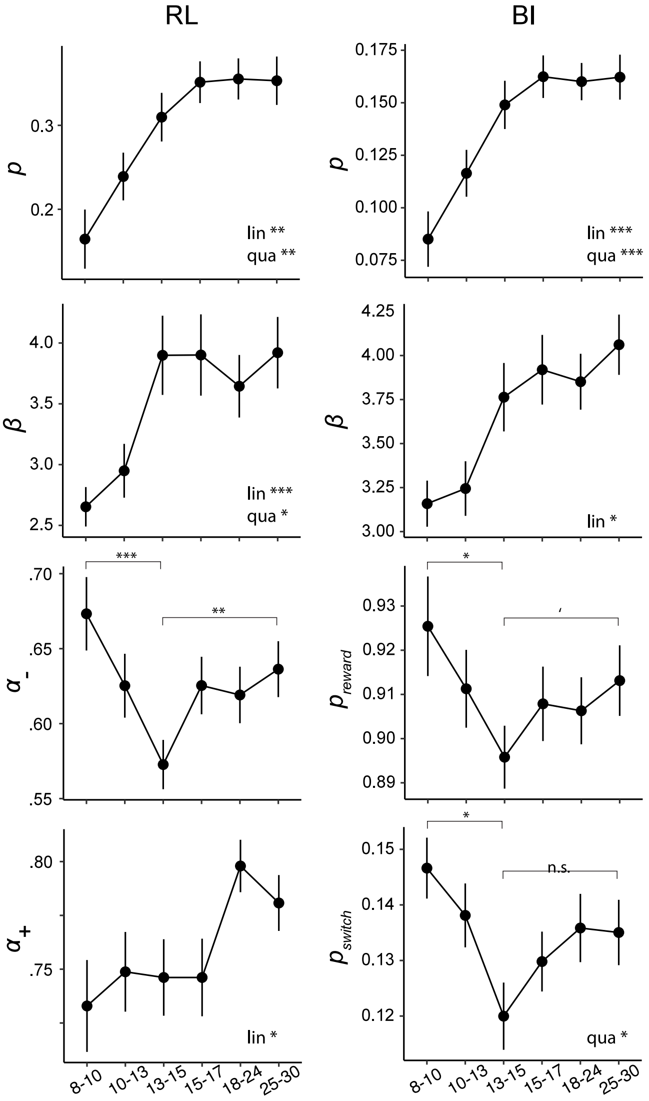

Fig. 5. Fitted model parameters for the winning RL (left column) and BI model (right), plotted over age. Stars in combination with "lin" or "qua" indicate significant linear ("lin") and quadratic ("qua") effects of age on model parameters, based on the age-based fitting model (Section 4.5.3). Stars on top of brackets show differences between groups, as revealed by t-tests conducted within the "age-less fitting model" (Section 4.5.3; suppl. Tables 13 and 14). Dots (means) and error bars (standard errors) show the results of the age-less fitting model, providing an unbiased representation of individual fits.

2.4. Integrating RL and BI—Going beyond specific models

These results raise an important question: Given that both RL and BI fit human behavior well, how do we reconcile differences in their computational mechanisms? To address this, we first determined whether both models covertly employed similar computational processes, predicting the same behavior despite differences in form. A generate-and-recover analysis, however, confirmed that they truly employed different processes (Wilson and Collins, 2019; Heathcote et al., 2015; Appendix 6.3.7).

We next asked whether both models captured similar aspects of cognition by assessing whether parameters were correlated between models. Parameters p and β were almost perfectly correlated, suggesting high consistency between models when estimating choice processes (for regression coefficients, see Fig. 6D). In addition, parameter p_{reward} (BI) was strongly correlated with α_- (RL), suggesting that beliefs about task stochasticity and learning rates for negative outcomes played

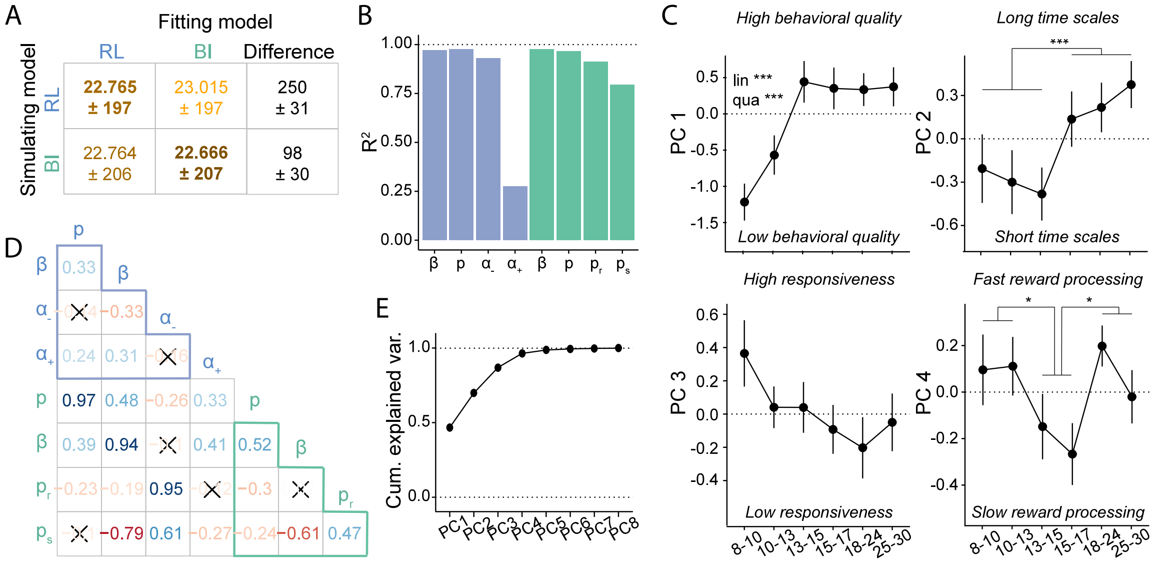

Fig. 6. Relating RL and BI models. (A) Model recovery. WAIC scores were worse (larger; lighter colors) when recovering behavior that was simulated from one model (row) using the other model (column), than when using the same model (diagonal), revealing that the models were discriminable. The difference in fit was smaller for BI simulations (bottom row), suggesting that the RL model captured BI behavior better than the other way around (top row). (B) Variance of each parameter explained by parameters and interactions of the other model ("R^2"), estimated through linear regression. All four BI parameters (green) were predicted almost perfectly by the RL parameters, and all RL parameters except for α_+ (RL) were predicted by the BI parameters. (D) Spearman pairwise correlations between model parameters. Red (blue) hue indicates negative (positive) correlation, saturation indicates correlation strength. Non-significant correlations are crossed out (Bonferroni-corrected at $p = 0.00089$). Light-blue (teal) letters refer to RL (BI) model parameters. Light-blue/teal-colored triangles show correlations within each model, remaining cells show correlations between models. (C) & (E) Results of PCA on model parameters (Section 4.5.5). (C) Age-related differences in PC1–4: As revealed by PC-based model simulations (Appendix 6.3.11), PC1 reflected overall behavioral quality. It showed rapid development between ages 8–13, which were captured by linear ("lin") and quadratic ("qua") effects in a regression model. PC2 captured a step-like transition from shorter to longer updating time scales at age 15. PC3 showed no significant age effects. PC4 captured the variance in α_+ and differed between adolescents 15–17 and both 8–13 year olds and adults. PC2 and PC4 were analyzed using t-tests. * $p < .05$; ** $p < .01$, *** $p < .001$. (E) Cumulative variance explained by all principal components PC1–8. The first four components analyzed in more detail captured 96.5% of total parameter variance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

similar roles across models, presumably in participants' response to negative outcomes. The other mental-model parameter, p_{switch} (BI), was strongly negatively correlated with β (RL), suggesting that beliefs about task volatility in the BI model captured aspects that were explained by decision noise in the RL model. This is consistent with the observation that an agent that expects high volatility could be mistaken for one that acts very noisily, given that both will make choices that are inconsistent with previous outcomes. The only parameter that showed no large correlations with other parameters was α_+ (RL), potentially reflecting a cognitive process uniquely captured by RL. Taken together, some parameters likely captured similar cognitive processes in both models, despite differences in their functional form, shown by large correlations between models. Other parameters were more unique, potentially reflecting model-specific cognitive processes. Further analyzes confirmed high shared explained variance between both models, using multiple regression (Appendix 6.3.9).

So far, we have provided two separate cognitive explanations for why adolescents performed our task better than other age groups: RL poses differences in value learning as the main driver, whereas BI poses differences in mental model-based inference. Could a single, broader explanation combine these insights and provide more general understanding of adolescent cognitive processing? To test this, we used PCA to unveil the lower-dimensional structure embedded in the 8-dimensional parameter space created by both models (Section 4.5.5). We found that the PCA's first four principle components (PCs) explained almost all variance (96.5%; Fig. 6E), showing that individual differences in all 8 model parameters could be summarized by just 4 abstract cognitive dimensions, which distill the insights of both models while abstracting away redundancies. To understand what these abstract dimensions reflected, we directly assessed which effects each PC had on behavior by simulating behaviors at high and low values of each PC (taking advantage of the fact that PCs were linear combinations of original model parameters, and could therefore be used directly to simulate behavior using our models; Appendix 6.3.11; Table 16).

This analysis revealed that PC1, capturing the largest proportion of parameter variance, reflected a broad measure of behavioral quality: Varying PC1 in simulation had striking effects on performance, leading to low-accuracy, random choices at small values of PC1 but optimal, highly accurate behavior at large values (suppl. Fig. 21A). PC2 represented integration time scales: At small values, simulated choices were only based on one previous trial, whereas at large values, several past trials mattered (suppl. Fig. 21B). PC3 captured responsiveness to task outcomes: At low values, recent outcomes only affected behavior minimally, whereas at high values, simulations seemed to overreact to the most recent outcome (suppl. Fig. 21C). PC4 uniquely captured RL parameter α_+, which was the only parameter with a non-negligible weight on this PC (suppl. Fig. 21D). Appendix 6.3.11 provides a more detailed description of each PC.

Three of these four PCs (PC1, PC2, PC4) showed prominent age effects: PC1 (behavioral quality) increased drastically until age 13, at which it reached a stable plateau that lasted—unchanged—throughout adulthood (Fig. 6C, top-left). Regression models revealed significant linear and quadratic effects of age on PC1 (lin.: $\beta = -0.47$, $t = -4.0$, $p < 0.001$; quad.: $\beta = 0.011$, $t = 3.43$, $p < 0.001$), with no effect of sex ($\beta = 0.020$, $t = 0.091$, $p = 0.93$). This suggests that the left side of the U-shaped trajectory in task performance (Fig. 2; Fig. 3C–F) might be caused by the development of behavioral quality (PC1): The peak in 13-to-15-year-olds compared to younger participants could be explained by the fact that 13-to-15-year-olds had already reached adult levels of behavioral quality, while younger participants showed noisier, less focused, and less consistent behavior.

By contrast, PC2 (updating time scales) followed a step function, such that participants in the three youngest age bins (8–15 years) acted on shorter times scales than participants in the three oldest bins (15–30; Fig. 6C, top-right; post-hoc t-test comparing both groups: $t(266.2) =$

3.44, $p < 0.001$). This pattern is in accordance with the interpretation that children's shorter time scales, facilitating rapid behavioral switches (suppl. Fig. 21B, left), were more beneficial for the current task than adults' longer time scales, which impeded switching (suppl. Fig. 21B, right). Differences in subjective time scale might therefore be the determining factor that allowed adolescents to outperform older participants, including adults.

PC4 (positive updates) differentiated the two adolescent age bins (13–17) from both younger (8–13) and older (18–30) participants (Fig. 6C, bottom-right), as revealed by significant post-hoc, Bonferroni-corrected, t-tests (8–13 vs. 13–17: $t(176.8) = 2.28$, $p = 0.047$; 13–17 vs. 18–30: $t(176.6) = 2.49$, $p = 0.028$). In other words, after accounting for variance in PC1-PC3, the remaining variance was explained by 13-to-17-year-olds' relatively longer updating timescales for positive outcomes (positive outcomes had relatively weaker immediate, but stronger long-lasting effects).

In sum, the PCA revealed four dimensions that combined the findings of both computational models, allowing for more abstract insights into developmental cognitive differences: Adolescents' unique competence in our task might be the result of adult-like behavioral quality in combination with child-like time scales and unique adolescent processing of positive feedback.

3. Discussion

Here, we tested performance in a volatile and stochastic reversal task across adolescent development in a large sample age 8–30. Several behavioral measures—including overall accuracy, number of points won, number of blocks completed, staying on potential switch trials, and asymptotic performance—suggested a mid-adolescent peak in performance. Our data suggest that the mid-adolescent age group recruited specific behavioral strategies to achieve this peak, including ignoring non-diagnostic negative feedback (Fig. 2E), showing persistent choices during stable task periods (Fig. 2F), and using negative feedback near-optimally (suppl. Fig. 8B, C, E, F). We statistically tested for adolescent peaks using a variety of methods, including two-line regression to prove the existence of U-shapes (Table 1; suppl. Fig. 9), quadratic regression to verify non-linear developments (Table 2), rolling averages to map the developmental trajectories more precisely (Fig. 2; suppl. Fig. 8A–C), pairwise t-tests to ascertain age group differences (Appendix 6.3.4), and age-bin analyzes to obtain the age at peak (Fig. 3C–F).

Though the precise outcomes varied slightly between different tests and behavioral measures, all supported a non-linear development of performance on the current task, and specifically the existence of a mid-adolescent performance peak. The results are also consistent with the idea that several cognitive functions contributed to this peak, including such with linear, saturating, and peaking patterns, explaining the mix of linear, quadratic, and U-shaped results. Similarly, previous research on a closely related task has suggested a combination of linear and U-shaped cognitive developments (van der Schaaf et al., 2011).

To test this prediction, we next investigated the cognitive processes that underlie this performance advantage. We considered that adolescents might integrate information over fewer trials than younger or older participants, as suggested, e.g., by (Davidow et al., 2016), or that they might process particular feedback types differently (e.g., (Palminteri et al., 2016)). These hypotheses can be tested using computational modeling in the RL framework, which explicitly estimates learning rates, and can reveal a differentiation between feedback types.

It is also possible, however, that adolescents outperformed other participants due to a better understanding of the task dynamics, which would allow them to predict more accurately when a switch occurred. Indeed, others have argued that both "model-based" behavior (Decker et al., 2016) and the tendency to employ counterfactual reasoning (Palminteri et al., 2016) improve with age, suggesting potential age differences in the quality of mental task models. The BI framework is ideal for testing this hypothesis because it explicitly targets participants' mental models and inference processes.

(However, it should be noted that the RL model does not lack the ability to draw inferences—the crucial difference is that inference is not made explicit in the RL model, just as learning is not made explicit in the BI model, even though it does not lack the ability to learn; see Appendix 6.4.3).

Another potential cognitive explanation is that adolescents might explore differently (Gopnik et al., 2017; Lloyd et al., 2020; Somerville et al., 2017), or be more persistent, a behavioral pattern commonly linked to PFC function (Morris et al., 2016; Kehagia et al., 2010), which continues maturation during adolescence (DePasque and Galván, 2017; Giedd et al., 1999; Toga et al., 2006). Whereas the previous two hypotheses targeted the "updating" step of decision making, these hypotheses concern the "choice" step, which can be tested in both RL and BI frameworks.

Our computational models revealed that several cognitive explanations are possible for our behavioral results: The RL model showed reduced learning speeds for negative outcomes in mid-adolescence (Fig. 5, left), supporting developmental differences in feedback processing and a differentiation of feedback types. The BI model suggested improved mental model parameters, supporting developmental differences in mental models and inference (Fig. 5, right). Crucially, the quantitative fit of both models to human data was similar (Table 3), and they both qualitatively reproduced human behavior in simulation (Fig. 3; Fig. 4), suggesting that both explanations are valid.

Furthermore, both models agreed on developmental differences in exploration/exploitation and persistence, conferring to the last hypothesis: Both showed monotonic trajectories between childhood and adulthood (Fig. 5, top two rows), which might support or modulate observed U-shaped patterns. Taken together, our study suggests that the observed adolescent performance advantage in a stochastic and volatile environment can be explained both by more adaptive negative feedback processing and by more optimal mental models.

Both explanations, however, are framed within a specific computational model. Can we draw more general conclusions? A common method to combine the insights of multiple models is to create a mixture model—we decided against this possibility for several reasons (Appendix 6.4.3), and instead used PCA to achieve this goal. This analysis revealed that developmental changes can be captured by three abstract, model-independent dimensions that vary with age: behavioral quality (PC1), time scales (PC2), and reward processing (PC4). Behavioral quality—likely a result of understanding of the task and experimental context, participant compliance, and attentional focus—showed a saturating pattern, reaching adult levels around the mid-adolescent age of peak performance, with no later age-related differences. Time scales, on the other hand—likely capturing an extended planning horizon, long-term credit assignment, memory, and prolonged attention—only increased after the performance peak, in late adolescence (Appendix 6.3.1). Finally, reward processing was slower at the age of the observed peak compared to both younger or older participants. Taken together, adolescents' behavioral advantage might be a combination of already adult-like quality of behavior, still child-like learning time scales, and unique reward processing.

3.1. Setting or adaptation?

These findings can be interpreted in two ways (Nussenbaum and Hartley, 2019): (1) Based on a *settings* account, these patterns are developmentally fixed, i.e., expected to guide behavior across contexts (including experiments and real-life situations). Specifically, our results would suggest that adolescents generally integrate negative feedback more slowly than other age groups, and generally expect fewer rewards (p_{reward}) and less volatility (p_{switch}). (2) The *adaptation* account, on the other hand, states that experimental findings are a function of both context (i.e., experimental task) and participants, and specific parameter values reveal the adaptability of participants to contexts, rather than universal behavioral tendencies. In this view, our results would highlight adolescents' increased adaptability to volatile and stochastic environments, given their ability to select near-optimal parameter settings for this task.

A recent review (Nussenbaum and Hartley, 2019) showed favorable empirical evidence for the adaptation compared to the settings account, given that specific parameter results often differ widely between studies, while parameter adaptiveness tends to be consistent (also see (Eckstein et al., 2021b,a)). Our results, as well, are consistent based on an adaption account, but contradict previous research based on a settings account: In a previous study (van der Schaaf et al., 2011), adolescents responded in the most *balanced* way to reward and punishment (Fig. 3A; children and adults responded more strongly to punishment and rewards, respectively); in our study, however, they responded in the most *imbalanced* way, responding least strongly to negative feedback. While these results contradict each other based on a settings view (greatest balance versus greatest imbalance), they both suggest that adolescents adapted best to the specific task demands, supporting an adaptation-based view: In (van der Schaaf et al., 2011), both positive and negative outcomes were diagnostic, requiring balanced learning, whereas in our study, only positive outcomes were diagnostic, requiring imbalanced learning. In other words, both studies suggest that adolescents showed an increased ability to quickly and effortlessly adapt various cognitive parameters to specific task demands when faced with stochastic, volatile contexts.

3.2. General cognitive abilities

A caveat of our study is the use of a cross-sectional rather than longitudinal design. We cannot exclude, for example, that adolescents had better schooling, a higher socio-economic status, or higher IQ scores than participants of other ages. If this was the case, the performance peak in adolescence might reflect a difference in task-unrelated factors rather than unique adaptation to stochasticity and volatility. However, several arguments speak against this possibility, including recruitment procedures, supplementary analyzes, and the distinctness of the U-shaped pattern observed in this task compared to the linear trajectories observed in other tasks performed by the same sample (Appendix 6.4.2).

3.3. A role of puberty?

Despite showing specific age-related differences, our study does not elucidate which biological mechanisms underlie these differences. There is growing evidence that gonadal hormones affect inhibitory neurotransmission, spine pruning, and other variables in the prefrontal cortex of rodents (Delevich et al., 2019, 2018; Juraska and Willing, 2017; Piekarski et al., 2017a,b; Drzewiecki et al., 2016), and evidence for puberty-related neurobehavioral change is also accumulating in human studies (Gracia-Tabuenca et al., 2021; Laube et al., 2020b; Op de Macks et al., 2016; Braams et al., 2015; Blakemore et al., 2010). To test if gonadal hormones might play a role in some of the observed differences, we assessed pubertal status through self-report questionnaires on physical maturation (Petersen et al., 1988) and salivary testosterone levels (for details, see (Master et al., 2020)). While some trends emerged with regard to early puberty (Appendix 6.3.5), our study was inconclusive on this issue. The observed trends warrant deeper investigation using longitudinal designs (Kraemer et al., 2000).

3.4. Dual-model approach to cognitive modeling

Because *basic* RL and BI models (Section 1.2) differ in their cognitive mechanisms (Sections 4.5.1 and 4.5.2) and behavior (suppl. Fig. 18 and Fig. 19), both complement each other: Each makes unique predictions, based on unique mechanisms, such that both jointly explain more than each would individually. However, in the current study, we "augmented" both models to approximate humans more closely (i.e., splitting learning rates, adding perseverance), thereby rendering their behavior and computational mechanisms more similar to each other. Does this increased similarity pose a problem for their joint use?

At least two arguments justify their combination: (1) Each model *explains* the cognitive process differently. Whereas RL explains it in terms of learning and differentiation of outcome types, BI explains it in terms of mental-model based predictions and inference. Hence, invoking different cognitive concepts, both explanations are non-redundant and provide additive insights. (2) Both models still *differ* behaviorally (Fig. 6A; suppl. Fig. 20; Appendix 6.3.8) and in terms of the cognitive processes captured by model parameters (Fig. 6B and D). This implies that both models captured different aspects of human cognitive processing and provided additive insights.

Taking a step back, the most common computational modeling approach selects just one family of candidate models (e.g., RL) and identifies the best-fitting one within this family, readily interpreting it as the cognitive process employed by participants. An issue with this approach is that a model from a different family (e.g., BI) might provide a better fit than any of the tested models. This issue can only be addressed by fitting models of multiple families, ensuring better coverage of the space of cognitive hypotheses.

However, this approach poses the new challenge that in addition to quantitative criteria of model fit (e.g., behavioral prediction, complexity; Bayes factor, AIC; Mulder and Wagenmakers, 2016; Pitt and Myung, 2002; Watanabe, 2013), qualitative criteria become increasingly important (e.g., interpretability, appropriateness for current hypotheses, conciseness, generality; Kording et al., 2020; Uttal, 1990; Webb, 2001; Blohm et al., 2020). Qualitative criteria are often more difficult to judge because they depend on scientific goals (e.g., explanation versus prediction; Navarro, 2019; Bernardo and Smith, 2009) and research philosophy (Blohm et al., 2020). Furthermore, qualitative and quantitative criteria can be at odds, inconveniencing model selection (Jacobs and Grainger, 1994).

To alleviate these issues, we focused on a range of criteria, including numerical fit (WAIC; slight advantage for RL), reproduction of participant behavior (equally good), continuity with previous neuroscientific research (RL), link to specific neural pathways (RL), centrality for developmental research (equal), claimed superiority in current paradigm (BI), and interpretability (BI: model parameters map directly onto main concepts p_{switch}: stochasticity, p_{reward}: volatility) to select a model. Because no model was obviously inferior to the other one based on these criteria, and both fitted behavior equally well, we opted to select two winners. This provided the benefits of *converging evidence* (e.g., replication: $\beta_{RL} \leftrightarrow \beta_{BI}$, $p_{RL} \leftrightarrow p_{BI}$; parallelism between models: $p_{reward} \leftrightarrow \alpha_{-}$), *distinct insights* (e.g., RL: importance of learning, differential processing of feedback types; BI: importance of inference, mental models), and the possibility to *combine* both models to expose more abstract factors (PC1, PC2, PC4) that differentiate adolescent cognitive processing from younger and older participants.

However, future research will be required to investigate these issues in more detail. It will be especially important to investigate the implications of obtaining evidence for more than one, fundamentally different, computational claims (e.g., RL and BI). Furthermore, new ways need to be outlined for dealing with ambiguous model comparison results (e.g., similar quality of simulated behavior and similar quantitative model fit, despite theoretically successful model recovery). We expect that issues like these will gain in prominence as more researchers adopt computational methods, and as the variety and quality of computational models increases. It is possible that specifically-designed experiments will be able to arbitrate between different types of computations; however, it is also possible that they will simply open new questions, for example showing joint contributions of different computations, or a more fundamental failure of all candidate computations. Likely, current experimentation practices (e.g., limited sample sizes, simplistic tasks) will also need to be revised to address these issues, and our toolkit of model fitting techniques might have to be equipped with new conceptual and methodological tools. While the current paper provides one way of dealing with a situation where no simple arbitration is possible, it does not offer clear-cut conclusions. We hope that, with more awareness of the theoretical issues around model comparison brought forth here, future research will address these questions more satisfactorily.

3.5. Conclusion

In conclusion, we showed that adolescents outperformed younger participants and adults in a volatile and stochastic context, two factors that were hypothesized to have specific relevance to the adolescent transition to independence. We used two computational models to examine the cognitive processes underlying this development, RL and BI. These models suggested that adolescents achieved better performance for different reasons: (1) They were best at accurately assessing the volatility and stochasticity of the environment, and integrated negative outcomes most appropriately (U-shapes in p_{reward}, p_{switch}, and α_{-}). (2) They combined adult-like behavioral quality (PC1), child-like time scales (PC2), and developmentally-unique processing of positive outcomes (PC4). Pubertal development and steroid hormones may impact a subset of these processes, yet causality is difficult to determine without manipulation or longitudinal designs (Kraemer et al., 2000).

For purposes of translation from the lab to the “real world”, our study indicates that how youth learn and decide changes in a nonlinear fashion as they grow. This underscores the importance of youth-serving programs that are developmentally informed and avoid a one-size-fits-all approach. Finally, these data support a positive view of adolescence and the idea that the adolescent brain exhibits remarkable learning capacities that should be celebrated.

4. Methods

4.1. Participants

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, using online ads (e.g., on neighborhood forums), flyers at community events (e.g., local farmers markets), and physicals posts in the neighborhood (e.g., printed ads). Community participants completed a battery of computerized tasks, questionnaires, and saliva samples (Master et al., 2020). In addition, 66 university undergraduate students (aged 18–50) were recruited through UC Berkeley’s Research Participation Pool, and completed the same four tasks, but not the pubertal-development questionnaire (PDS; Petersen et al., 1988) or saliva sample. Community participants were prescreened for the absence of present or past psychological and neurological disorders; the undergraduate sample indicated the absence of these. Community participants were compensated with 25$ for the 1–2 h in-lab portion of the experiment and 25$ for completing optional take-home saliva samples; undergraduate students received course credit for participation.

Exclusion criteria. Out of the 191 participants under 18, 184 completed the current task; reasons for not completing the task included getting tired, running out of time, and technical issues. Five participants (mean age 10.0 years) were excluded because their mean accuracy was below 58% (chance: 50%), an elbow point in accuracy, which suggests they did not pay attention to the task. This led to a sample of 179 participants under 18 (male: 96, female: 83). Two participants from the undergraduate sample were excluded because they were older than 30, leading to a sample aged 18–28; 7 were excluded because they failed to indicate their age. This led to a final sample of 57 undergraduate participants (male: 19, female: 38). All 55 adult community participants (male: 26, female: 29) completed the task and were included in the analyzes, leading to a sample size of 179 participants below 18, and 291 in total (suppl. Fig. 11).

4.2. Testing procedure

After entering the testing room, participants under 18 years and their guardians provided informed assent and permission, respectively; 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 (for more details about the other tasks, see (Master et al., 2020; Xia et al., 2020; Eckstein et al., 2021a); for a comparison of all tasks, see (Eckstein et al., 2021b,a)). At the conclusion of the tasks, participants between 11 and 18 completed the PDS questionnaire, and all participants were measured in height and weight and compensated with $25 Amazon gift cards. The entire session took 2–3 h for community participants (e.g., some younger participants took more breaks), and 1 h for undergraduate participants (who did not complete the puberty measures and saliva sample). We paid great attention to the fact that participants took sufficient breaks between tasks to avoid excessive fatigue and limit the effects of the differences in testing duration.

4.3. Task design

The goal of the task was to collect golden coins, which were hidden in one of two boxes. On each trial, participants decided which box to open, and either received a reward (coin) or not (empty). Task contingencies—i.e., which box was correct and therefore able to produce coins—switched unpredictably throughout the task (Fig. 1B). Before the main task, participants completed a 3-step tutorial: (1) A prompt explained that only one of the boxes contained a coin (was “magical”), and participants completed 10 practice trials on which one box was always rewarded and the other never (deterministic phase). (2) Another prompt explained that the magical box sometimes switches sides, and participants received 8 trials on which only second box was rewarded, followed by 8 trials on which only the first box was rewarded (switching phase). (3) The last prompt explained that the magical box did not always contain a coin, and led into the main task with 120 trials.

In the main task, the correct box was rewarded in 75% of trials; the incorrect box was never rewarded. After participants reached a performance criterion, it became possible for contingencies to switch (without notice), such that the previously incorrect box became the correct one. The performance criterion was to collect 7–15 rewards, with the specific number pre-randomized for each block (any number of non-rewarded trials was allowed in-between rewarded trials). Switches only occurred after rewarded trials, and the first correct choice after a switch was always rewarded (while retaining an average of 75% probability of reward for correct choices), for consistency with the rodent task (Tai et al., 2012).

4.4. Behavioral analyzes

For the "countable" performance measures "number of points won" and "number of blocks completed", we calculated corrected measures because—after excluding invalid trials—some participants had fewer trials than the original 120. For both measures, corrected measures m were calculated based on the raw counts r of each measure in the final dataset (with valid number of trails t), as follows: $m = 120 * \frac{r}{t}$.

We calculated age-based rolling performance averages by averaging the mean performance of 50 subsequent participants ordered by age. Standard errors were calculated based on the same rolling window.

We assessed the effects of age on behavioral outcomes (Fig. 2), using (logistic) mixed-effects regression models using the package lme4 (Bates et al., 2015) in R (RCoreTeam, 2016). All models included the following regressors to predict outcomes (e.g., overall accuracy, response times): Z-scored age, to assess the linear effect of age on the outcome; squared, z-scored age, to assess the quadratic (U-shaped) effect of age; and sex; furthermore, all models specified random effects of participants, allowing participants' intercepts and slopes to vary independently. Additional predictors are noted in the main text. For example, the formula of the overall accuracy model was: `glmer(ACC ~ age_z + age_z_squared + sex + (1 | participant), data, binomial)`, where `data` refers to the trial-wise behavioral data of each participant.

We assessed the effects of previous outcomes on participants' choices (suppl. Fig. 8B, C, E, F) using logistic mixed-effects regression, predicting actions (left, right) from previous outcomes (details below), while testing for effects of and interactions with sex, z-scored age, and z-scored quadratic age, specifying participants as mixed effects. We included one predictor for positive and one for negative outcomes at each delay i with respect to the predicted action (e.g., $i = 1$ trial ago). Outcome predictors were coded −1 for left and +1 for right choices (0 otherwise). Including predictors of trials $1 \leq i \leq 8$ provided the best model fit (suppl. Table 8): $AIC_{i \leq 3}$: 31.046; $AIC_{i \leq 4}$: 31.013; $AIC_{i \leq 5}$: 31.001; $AIC_{i \leq 6}$: 30.981; $AIC_{i \leq 7}$: 30.963; $AIC_{i \leq 8}$: **30.962**; $AIC_{i \leq 9}$: 30.966; $AIC_{i \leq 10}$: 30.964. To visualize the results of this model including all participants, we also ran separate models for each participant (suppl. Fig. 8B, C, E, F). However, due to issues of multiple comparisons, the grand model was used to assess statistical significance.

We conducted the two-lines regression models according to the method specified in (Simonsohn, 2018), using the functions provided by the author at http://webstimate.org/twolines/.

4.5. Computational models

4.5.1. Reinforcement Learning (RL) models

A basic RL model has two parameters, learning rate α and decision temperature β. On each trial t, the value $Q_t(a)$ of action a is updated based on the observed outcome $o_t \in [0, 1]$ (no reward, reward):

$$Q_{t+1}(a) = Q_t(a) + \alpha(o_t - Q_t(a))$$

Action values inform choices probabilistically, based on a softmax transformation:

$$p_t(a) = \frac{exp(\beta\ Q_t(a))}{exp(\beta\ Q_t(a)) + exp(\beta\ Q_t(a_{ns}))}$$

Here, a is the selected, and a_{ns} the non-selected action.

Compared to this basic 2-parameter model, the best-fit 4-parameter model was augmented by splitting learning rates into α_+ and α_-, adding persistence parameter p, and the ability for counterfactual updating. We explain each in turn: Splitting learning rates allowed to differentiate updates for rewarded ($o_t = 1$) versus non-rewarded ($o_t = 0$) trials, with independent α_- and α_+:

$$Q_{t+1}(a) = \begin{cases} Q_t(a) + \alpha_+(o_t - Q_t(a)), & \text{if } o_t = 1 \\ Q_t(a) + \alpha_-(o_t - Q_t(a)), & \text{if } o_t = 0 \end{cases}$$

Choice persistence or "stickiness" p changed the value of the previously-selected action a_t on the subsequent trial, biasing toward staying ($p > 0$) or switching ($p < 0$):

$$Q_{t+1}(a) = \begin{cases} Q_{t+1}(a) + p, & \text{if } a_t = a_{t-1} \\ Q_{t+1}(a), & \text{if } a_t \neq a_{t-1} \end{cases}$$

Counterfactual updating allows updates to non-selected actions based on counterfactual outcomes $1 - o_t$:

$$Q_{t+1}(a_{ns}) = \begin{cases} Q_t(a_{ns}) + \alpha_+((1 - o_t) - Q_t(a_{ns})), & \text{if } o = 1 \\ Q_t(a_{ns}) + \alpha_-((1 - o_t) - Q_t(a_{ns})), & \text{if } o = 0 \end{cases}$$

Initially, we used four parameters α_+, α_{+c}, α_-, and α_{-c} to represent each combination of value-based ("+" versus "−") and counter-factual ("c") versus factual updating, but collapsing $\alpha_+ = \alpha_{+c}$ and $\alpha_- = \alpha_{-c}$ improved model fit (Table 3). This suggests that outcomes triggered equal-sized updates to chosen and unchosen actions.

This final model can be interpreted as basing decisions on a single value estimate (value difference between both actions), rather than independent value estimates for each action because chosen and unchosen actions were updated to the same degree and in opposite directions on each trial. Action values were initialized at 0.5 for all models.

4.5.2. BayesIan Inference (BI) models

The BI model is based on two hidden states: "Left action is correct" ($a_{left} = cor$) and "Right action is correct" ($a_{right} = cor$). On each trial, the hidden state switches with probability p_{switch}. In each state, the probability of receiving a reward for the correct action is p_{reward} (Fig. 3A). On each trial, actions are selected in two phases, using a Bayesian Filter algorithm (Sarkka, 2013): (1) In the *estimation phase*, the hidden state of the previous trial $t - 1$ is inferred based on outcome o_{t-1}, using Bayes rule:

$$p(a_{t-1} = cor \mid o_{t-1}) = \frac{p(o_{t-1}|a_{t-1} = cor)\ p(a_{t-1} = cor)}{p(o_{t-1}|a_{t-1} = cor)\ p(a_{t-1} = cor) + p(o_{t-1}|a_{t-1} = inc)\ p(a_{t-1} = inc)}$$

$p(a_{t-1} = cor)$ is the prior probability that a_{t-1} is correct (on the first trial, $p(a = cor) = 0.5$ for a_{left} and a_{right}). $p(o_{t-1}|a_{t-1})$ is the likelihood of the observed outcome o_{t-1} given action a_{t-1}. Likelihoods are (dropping underscripts for clarity): $p(o = 1|a = cor) = p_{reward}$, $p(o = 0|a = cor) = 1 - p_{reward}$, $p(o = 1|a = inc) = \epsilon$, and $p(o = 0|a = cor) = 1 - \epsilon$. ϵ is the probability of receiving a reward for an incorrect action, which was 0 in reality, but set to $\epsilon = 0.0001$ to avoid model degeneracy.

(2) In the *prediction phase*, the possibility of state switches is taken into account by propagating the inferred hidden-state belief at $t - 1$ forward to trial t:

$$p(a_t = cor) = (1 - p_{switch})\ p(a_{t-1} = cor) + p_{switch}\ p(a_{t-1} = inc)$$

We first assessed a parameter-free version of the BI model, truthfully setting $p_{reward} = 0.75$, and $p_{switch} = 0.05$. Lacking free parameters, this model was unable to capture individual differences and led to poor qualitative (suppl. Fig. 19A) and quantitative model fit (Table 3). The best-fit BI model had four free parameters: p_{reward} and p_{switch}, as well as the choice parameters β and p, like the winning RL model. β and p were introduced by applying a softmax to $p(a_t = cor)$ to calculate $p(a_t)$, the probability of selecting action a on trial t:

$$p(a_t) = \frac{1}{1 + exp(\beta(0.5 - p - p(a_t = cor)))}$$

When both actions had the same probability and persistence $p > 0$, then staying was more likely; when $p < 0$, then switching was more likely.

4.5.3. Model fitting and comparison

We fitted parameters using hierarchical Bayesian methods (Lee, 2011; Katahira, 2016; van den Bos et al., 2017; Fig. 3B), whose parameter recovery clearly superseded those of classical maximum-likelihood fitting (suppl. Fig. 7). Rather than fitting individual participants, hierarchical Bayesian model fitting estimates the parameters of a population jointly by maximizing the posterior probability $p(\theta|data)$ of all parameters θ conditioned on the observed $data$, using Bayesian inference:

$$p(\theta|data) \propto p(data|\theta)\ p(\theta)$$

An advantage of hierarchical Bayesian model fitting is that individual parameters are embedded in a hierarchical structure of priors, which helps resolve uncertainty at the individual level.

We ran two models to fit parameters: The “age-less” model was used to estimate participants’ parameters in a non-biased way and conduct binned analyzes on parameter differences; the “age-based” model was used to statistically assess the shapes of parameters’ age trajectories. In the age-less model, each individual j’s parameters $\theta_j^{RL} = [p, \beta, \alpha_-, \alpha_+]$ or $\theta_j^{BI} = [p, \beta, p_{switch}, p_{reward}]$ were drawn from group-based prior parameter distributions. Parameters were drawn from appropriately-shaped prior distributions, limiting ranges where necessary, which where based on non-informative, appropriate hyper-priors (suppl. Table 6).

Next, we fitted the model by determining the group-level and individual parameters with the largest posterior probability under the behavioral data $p(\theta|data)$. Because $p(\theta|data)$ was analytically intractable, we approximated it using Markov-Chain Monte Carlo sampling, using the no-U-Turn sampler from the PyMC3 package in python (Salvatier et al., 2016). We ran 2 chains per model with 6,000 samples per chain, discarding the first 1,000 as burn-in. All models converged with small MC errors, sufficient effective sample sizes, and $\hat{R}$ close to 1 (suppl. Table 7). For model comparison, we used the Watanabe–Akaike information criterion (WAIC), which estimates the expected out-of-sample prediction error using a bias-corrected adjustment of within-sample error (Watanabe, 2013).

To obtain participants’ individual fitted parameters, we calculated the means over all posterior samples (Fig. 5, suppl. Figures 4, 18, and 19). To test whether a parameter θ differed between two age groups $a1$ and $a2$, we determined the number of MCMC samples in which the parameter was larger in one group than the other, i.e., the expectation $\mathbb{E}(\theta_{a1} < \theta_{a2})$ across MCMC samples. $p < 0.05$ was used to determine significance. This concludes our discussion of the age-less model, which was used to calculate individual parameters in an unbiased way.

To adequately assess the age trajectories of fitted parameters, we employed a fitting technique based on hierarchical Bayesian model fitting (Lee, 2011; Katahira, 2016), which avoids biases that arise when comparing parameters between participants that have been fitted using maximum-likelihood (van den Bos et al., 2017), and allows to test specific hypotheses about parameter trajectories by explicitly modeling these trajectories within the fitting framework: We conducted a separate “age-based” model, in which model parameters were allowed to depend on participants’ age (Fig. 3B). Estimating age effects directly within the computational model allowed us to estimate group-level effects in an unbiased way, whereas flat (hierarchical) models that estimate parameters but not age effects would underestimate (overestimate) group-level effects, respectively (Boehm et al., 2018). The age-based model was exclusively used to statistically assess parameter age trajectories because individual parameters would be biased by the inclusion of age in the model.

In the age-based model, each parameter θ of each participant j was sampled from a Normal distribution around an age-based regression line (Fig. 3B):

$$\theta_j \sim Normal(\mu = \theta_{int} + age\ \times \theta_{lin} + age^2\ \times \theta_{qua},\ \sigma = \theta_{sd})$$

Each parameter’s intercept θ_{int}, linear change with age θ_{lin}, quadratic change with age θ_{qua}, and standard deviation θ_{sd} were sampled from prior distributions of the form specified in suppl. Table 6. For more information, see supplementary section 6.2.2.

4.5.4. Correlations between model parameters (Fig. 6D)

We used Spearman correlation because parameters followed different, not necessarily normal, distributions. Employing Pearson correlation led to similar results. p-values were corrected for multiple comparisons using the Bonferroni method.

4.5.5. Principal Component Analysis (PCA)

To extract general cognitive components from model parameters, we ran a PCA on all fitted parameters (8 per participant). PCA can be understood as a method that rotates the initial coordinate system of a dataset (in our case, 8 axes corresponding to the 8 parameters), such that the first axis is aligned with the dimension of largest variation in the dataset (first principle component; PC1), the second axis with the dimension of second-largest variance (PC2), while being orthogonal to the first, and so on. In this way, all resulting PCs are orthogonal to each other, and explain subsequently less variance in the original dataset. We conducted a PCA after centering and scaling (z-scoring) the data, using R (RCoreTeam, 2016).

To assess PC age effects, we ran similar regression models as for behavioral measures, predicting PCs from z-scored age (linear), z-scored age (quadratic), and sex. When significant, effects were noted in Fig. 6C. For PC2 and PC4, we also conducted post-hoc t-tests, correcting for multiple comparison using the Bonferroni method (suppl. Table 17).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data are available on https://osf.io/7wuh4/ The research code has been shared on https://github.com/MariaEckstein/SLCN/blob/master/models/PSAllModels.pyPlease get in touch with any questions.

Acknowledgments

Numerous people contributed to this research: Amy Zou, Lance Kriegsfeld, Celia Ford, Jennifer Pfeifer, Megan Johnson, Gautam Agarwal, Liyu Xia, 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, USA SL-CN grant 1640885 to RD, AGEC, and LW.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.dcn.2022.101106.

References

Adleman, N., Kayser, R., Dickstein, D., Blair, R., Pine, D., Leibenluft, E., 2011. Neural correlates of reversal learning in severe mood dysregulation and pediatric bipolar disorder. J. Am. Acad. Child Adolesc. Psychiatry 50, 1173–1185.e2. http://dx.doi.org/10.1016/j.jaac.2011.07.011.

Albert, D., Chein, J., Steinberg, L., 2013. The teenage brain: Peer influences on adolescent decision making. Curr. Direct. Psychol. Sci. 22 (2), 114–120. http://dx.doi.org/10.1177/0963721412471347.

Bartolo, R., Averbeck, B.B., 2020. Prefrontal cortex predicts state switches during reversal learning. Neuron 106 (6), 1044–1054.e4. http://dx.doi.org/10.1016/j.neuron.2020.03.024.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Bernardo, J.M., Smith, A.F.M., 2009. Bayesian Theory. John Wiley & Sons, Google-Books-ID: 11nSgIcd7xQC.

Blakemore, S.-J., Burnett, S., Dahl, R.E., 2010. The role of puberty in the developing adolescent brain. Hum. Brain Mapp. 31 (6), 926–933. http://dx.doi.org/10.1002/hbm.21052.

Blakemore, S.J., Robbins, T.W., 2012. Decision-making in the adolescent brain. Nat. Neurosci. 15 (9), 1184–1191. http://dx.doi.org/10.1038/nn.3177, Number: 9 Publisher: Nature Publishing Group.

Blohm, G., Kording, K.P., Schrater, P.R., 2020. A how-to-model guide for neuroscience. eNeuro 7 (1), http://dx.doi.org/10.1523/ENEURO.0352-19.2019, Publisher: Society for Neuroscience Section: Research Article: Methods/New Tools.

Boehm, U., Marsman, M., Matzke, D., Wagenmakers, E.-J., 2018. On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. Behav. Res. Methods 50 (4), 1614–1631. http://dx.doi.org/10.3758/s13428-018-1054-3.

Boehme, R., Lorenz, R.C., Gleich, T., Romund, L., Pelz, P., Golde, S., Flemming, E., Wold, A., Deserno, L., Behr, J., Raufelder, D., Heinz, A., Beck, A., 2017. Reversal learning strategy in adolescence is associated with prefrontal cortex activation. Eur. J. Neurosci. 45 (1), 129–137. http://dx.doi.org/10.1111/ejn.13401.

Boorman, E.D., Behrens, T.E., Rushworth, M.F., 2011. Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex (M. L. Platt, Ed.). PLoS Biol. 9 (6), e1001093. http://dx.doi.org/10.1371/journal.pbio.1001093.

Braams, B.R., Duijvenvoorde, A.C.K., van and Peper, J.S., Crone, E.A., 2015. Longitudinal changes in adolescent risk-taking: A comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. J. Neurosci. 35 (18), 7226–7238. http://dx.doi.org/10.1523/JNEUROSCI.4764-14.2015.

Brandner, P., Güroğlu, B., van de Groep, S., Spaans, J.P., Crone, E.A., 2021. Happy for us not them: Differences in neural activation in a vicarious reward task between family and strangers during adolescent development. Dev. Cogn. Neurosci. 100985. http://dx.doi.org/10.1016/j.dcn.2021.100985.

Bromberg-Martin, E.S., Matsumoto, M., Hong, S., Hikosaka, O., 2010. A pallidus-habenula-dopamine pathway signals inferred stimulus values. J. Neurophysiol. 104 (2), 1068–1076. http://dx.doi.org/10.1152/jn.00158.2010, Publisher: American Physiological Society.

Casey, B.J., Jones, R.M., Hare, T.A., 2008. The adolescent brain. Ann. New York Acad. Sci. 1124 (1), 111–126. http://dx.doi.org/10.1196/annals.1440.010.

Cauffman, E., Shulman, E.P., Steinberg, L., Claus, E., Banich, M.T., Graham, S., Woolard, J., 2010. Age differences in affective decision making as indexed by performance on the iowa gambling task. Dev. Psychol. 46 (1), 193. http://dx.doi.org/10.1037/a0016128, Publisher: US: American Psychological Association.

Cazé, R.D., van der Meer, M.A.A., 2013. Adaptive properties of differential learning rates for positive and negative outcomes. Biol. Cybernet. 107 (6), 711–719. http://dx.doi.org/10.1007/s00422-013-0571-5.

Chase, H.W., Swainson, R., Durham, L., Benham, L., Cools, R., 2010. Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. J. Cogn. Neurosci. 23 (4), 936–946. http://dx.doi.org/10.1162/jocn.2010.21456.

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

Clark, L., Cools, R., Robbins, T.W., 2004. The neuropsychology of ventral prefrontal cortex: Decision-making and reversal learning. Brain Cogn. 55 (1), 41–53. http://dx.doi.org/10.1016/S0278-2626(03)00284-7.

Costa, V.D., Tran, V.L., Turchi, J., Averbeck, B.B., 2015. Reversal learning and dopamine: A Bayesian perspective. J. Neurosci. 35 (6), 2407–2416. http://dx.doi.org/10.1523/JNEUROSCI.1989-14.2015.

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

Dahl, R.E., Allen, N.B., Wilbrecht, L., Suleiman, A.B., 2018. Importance of investing in adolescence from a developmental science perspective. Nature 554 (7693), 441–450. http://dx.doi.org/10.1038/nature25770.

Davidow, J.Y., Foerde, K., Galvan, A., Shohamy, D., 2016. An upside to reward sensitivity: The hippocampus supports enhanced reinforcement learning in adolescence. Neuron 92 (1), 93–99. http://dx.doi.org/10.1016/j.neuron.2016.08.031.

Decker, J.H., Otto, A.R., Daw, N.D., Hartley, C.A., 2016. From creatures of habit to goal-directed learners. Psychol. Sci. 27 (6), 848–858. http://dx.doi.org/10.1177/0956797616639301.

Defoe, I.N., Dubas, J.S., Figner, B., van Aken, M.A.G., 2015. A meta-analysis on age differences in risky decision making: Adolescents versus children and adults. Psychol. Bull. 141 (1), 48–84. http://dx.doi.org/10.1037/a0038088.

Delevich, K., Piekarski, D., Wilbrecht, L., 2019. Neuroscience: Sex hormones at work in the neocortex. Curr. Biol. 29 (4), R122–R125. http://dx.doi.org/10.1016/j.cub.2019.01.013.

Delevich, K., Thomas, A.W., Wilbrecht, L., 2018. Adolescence and "late blooming" synapses of the prefrontal cortex. Cold Spring Harbor Symp. Quant. Biol. 83, 37–43. http://dx.doi.org/10.1101/sqb.2018.83.037507.

DePasque, S., Galván, A., 2017. Frontostriatal development and probabilistic reinforcement learning during adolescence. Neurobiol. Learn. Mem. 143, 1–7. http://dx.doi.org/10.1016/j.nlm.2017.04.009.

Dickstein, D.P., Finger, E.C., Brotman, M.A., Rich, B.A., Pine, D.S., Blair, J.R., Leibenluft, E., 2010a. Impaired probabilistic reversal learning in youths with mood and anxiety disorders. Psychol. Med. 40 (7), 1089–1100. http://dx.doi.org/10.1017/S0033291709991462.

Dickstein, D.P., Finger, E.C., Skup, M., Pine, D.S., Blair, J.R., Leibenluft, E., 2010b. Altered neural function in pediatric bipolar disorder during reversal learning. Bipolar Disorders 12 (7), 707–719. http://dx.doi.org/10.1111/j.1399-5618.2010.00863.x.

Drzewiecki, C.M., Willing, J., Juraska, J.M., 2016. Synaptic number changes in the medial prefrontal cortex across adolescence in male and female rats: A role for pubertal onset. Synapse 70 (9), 361–368. http://dx.doi.org/10.1002/syn.21909.

Eckstein, M.K., Master, S.L., Xia, L., Dahl, R.E., Wilbrecht, L., Collins, A.G.E., 2021a. Learning rates are not all the same: The interpretation of computational model parameters depends on the context. BioRxiv 2021.05.28.446162. http://dx.doi.org/10.1101/2021.05.28.446162, Publisher: Cold Spring Harbor Laboratory Section: New Results.

Eckstein, M.K., Wilbrecht, L., Collins, A.G., 2021b. What do reinforcement learning models measure? Interpreting model parameters in cognition and neuroscience. Curr. Opin. Behav. Sci. 41, 128–137. http://dx.doi.org/10.1016/j.cobeha.2021.06.004.

Finger, E.C., Marsh, A.A., Mitchell, D.G., Reid, M.E., Sims, C., Budhani, S., Kosson, D.S., Chen, G., Towbin, K.E., Leibenluft, E., Pine, D.S., Blair, J.R., 2008. Abnormal ventromedial prefrontal cortex function in children with psychopathic traits during reversal learning. Arch. Gen. Psychiatry 65 (5), 586–594. http://dx.doi.org/10.1001/archpsyc.65.5.586.

Frank, M.J., Claus, E.D., 2006. Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychol. Rev. 113 (2), 300–326. http://dx.doi.org/10.1037/0033-295X.113.2.300.

Frank, M.J., Seeberger, L.C., O'Reilly, R.C., 2004. By carrot or by stick: Cognitive reinforcement learning in parkinsonism. Science 306 (5703), 1940–1943. http://dx.doi.org/10.1126/science.1102941.

Frankenhuis, W.E., Walasek, N., 2020. Modeling the evolution of sensitive periods. Dev. Cogn. Neurosci. 41, 100715. http://dx.doi.org/10.1016/j.dcn.2019.100715.

Fuhs, M.C., Touretzky, D.S., 2007. Context learning in the rodent hippocampus. Neural Comput. 19 (12), 3173–3215. http://dx.doi.org/10.1162/neco.2007.19.12.3173.

Gershman, S.J., Uchida, N., 2019. Believing in dopamine. Nat. Rev. Neurosci. 20 (11), 703–714. http://dx.doi.org/10.1038/s41583-019-0220-7, Number: 11 Publisher: Nature Publishing Group.

Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L., 1999. Brain development during childhood and adolescence: a longitudinal MRI study. Nature Neurosci. 2 (10), 861–863. http://dx.doi.org/10.1038/13158.

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

Gopnik, A., O'Grady, S., Lucas, C.G., Griffiths, T.L., Wente, A., Bridgers, S., Aboody, R., Fung, H., Dahl, R.E., 2017. Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. Proc. Natl. Acad. Sci. 114 (30), 7892–7899. http://dx.doi.org/10.1073/pnas.1700811114.

Gracia-Tabuenca, Z., Moreno, M.B., Barrios, F.A., Alcauter, S., 2021. Development of the brain functional connectome follows puberty-dependent nonlinear trajectories. NeuroImage 229, 117769. http://dx.doi.org/10.1016/j.neuroimage.2021.117769.

Hamilton, D.A., Brigman, J.L., 2015. Behavioral flexibility in rats and mice: Contributions of distinct frontocortical regions. Genes Brain Behav. 14 (1), 4–21. http://dx.doi.org/10.1111/gbb.12191.

Harada, T., 2020. Learning from success or failure? – Positivity biases revisited. Front. Psychol. 11, http://dx.doi.org/10.3389/fpsyg.2020.01627.

Harden, K.P., Tucker-Drob, E.M., 2011. Individual differences in the development of sensation seeking and impulsivity during adolescence: further evidence for a dual systems model. Dev. Psychol. 47 (3), 739–746. http://dx.doi.org/10.1037/a0023279.

Harms, M.B., Bowen, K.E.S., Hanson, J.L., Pollak, S.D., 2018. Instrumental learning and cognitive flexibility processes are impaired in children exposed to early life stress. Dev. Sci. 21 (4), e12596. http://dx.doi.org/10.1111/desc.12596.

Hauser, T.U., Iannaccone, R., Ball, J., Mathys, C., Brandeis, D., Walitza, S., Brem, S., 2014. Role of the medial prefrontal cortex in impaired decision making in juvenile attention-deficit/hyperactivity disorder. JAMA Psychiatry 71 (10), 1165. http://dx.doi.org/10.1001/jamapsychiatry.2014.1093.

Hauser, T.U., Iannaccone, R., Walitza, S., Brandeis, D., Brem, S., 2015. Cognitive flexibility in adolescence: Neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. NeuroImage 104, 347–354. http://dx.doi.org/10.1016/j.neuroimage.2014.09.018.

Heathcote, A., Brown, S.D., Wagenmakers, E.-J., 2015. An introduction to good practices in cognitive modeling. In: Forstmann, B.U., Wagenmakers, E.-J. (Eds.), An Introduction to Model-Based Cognitive Neuroscience. Springer, New York, NY, pp. 25–48. http://dx.doi.org/10.1007/978-1-4939-2236-9_2, (B. U. Forstmann & E. -J. Wagenmakers, Eds.).

Hildebrandt, T., Schulz, K., Schiller, D., Heywood, A., Goodman, W., Sysko, R., 2018. Evidence of prefrontal hyperactivation to food-cue reversal learning in adolescents with anorexia nervosa. Behav. Res. Therapy 111, 36–43. http://dx.doi.org/10.1016/j.brat.2018.08.006.

Insel, C., Somerville, L.H., 2018. Asymmetric neural tracking of gain and loss magnitude during adolescence. Soc. Cogn. Affect. Neurosci. 13 (8), 785–796. http://dx.doi.org/10.1093/scan/nsy058.

Izquierdo, A., Brigman, J.L., Radke, A.K., Rudebeck, P.H., Holmes, A., 2017. The neural basis of reversal learning: An updated perspective. Neuroscience 345, 12–26. http://dx.doi.org/10.1016/j.neuroscience.2016.03.021.

Izquierdo, A., Jentsch, J.D., 2012. Reversal learning as a measure of impulsive and compulsive behavior in addictions. Psychopharmacology 219 (2), 607–620. http://dx.doi.org/10.1007/s00213-011-2579-7.

Jacobs, A.M., Grainger, J., 1994. Models of visual word recognition: Sampling the state of the art. J. Exp. Psychol. Hum. Percept. Perform. 20 (6), 1311–1334. http://dx.doi.org/10.1037/0096-1523.20.6.1311, Place: US Publisher: American Psychological Association.

Javadi, A.H., Schmidt, D.H.K., Smolka, M.N., 2014. Adolescents adapt more slowly than adults to varying reward contingencies. J. Cogn. Neurosci. 26 (12), 2670–2681.

Jepma, M., Schaaf, J.V., Visser, I., Huizenga, H.M., 2020. Uncertainty-driven regulation of learning and exploration in adolescents: A computational account. PLoS Comput. Biol. 16 (9), e1008276. http://dx.doi.org/10.1371/journal.pcbi.1008276, Publisher: Public Library of Science.

Johnson, C., Wilbrecht, L., 2011. Juvenile mice show greater flexibility in multiple choice reversal learning than adults. Dev. Cogn. Neurosci. 1 (4), 540–551. http://dx.doi.org/10.1016/j.dcn.2011.05.008.

Juraska, J.M., Willing, J., 2017. Pubertal onset as a critical transition for neural development and cognition. Brain Res. 1654 (Pt B), 87–94. http://dx.doi.org/10.1016/j.brainres.2016.04.012.

Katahira, K., 2016. How hierarchical models improve point estimates of model parameters at the individual level. J. Math. Psych. 73, 37–58. http://dx.doi.org/10.1016/j.jmp.2016.03.007.

Kehagia, A.A., Murray, G.K., Robbins, T.W., 2010. Learning and cognitive flexibility: frontostriatal function and monoaminergic modulation. Curr. Opin. Neurobiol. 20 (2), 199–204. http://dx.doi.org/10.1016/j.conb.2010.01.007.

Kleibeuker, S.W., Dreu, C.K.W.D., Crone, E.A., 2013. The development of creative cognition across adolescence: distinct trajectories for insight and divergent thinking. Dev. Sci. 16 (1), 2–12. http://dx.doi.org/10.1111/j.1467-7687.2012.01176.x.

Kording, K.P., Blohm, G., Schrater, P., Kay, K., 2020. Appreciating the variety of goals in computational neuroscience. Neurons Behav. Data Anal. Theory 3 (6), 1–12, Retrieved April 23, 2021, from https://nbdt.scholasticahq.com/article/16723-appreciating-the-variety-of-goals-in-computational-neuroscience Publisher: The neurons, behavior, data analysis and theory collective.

Kraemer, H.C., Yesavage, J.A., Taylor, J.L., Kupfer, D., 2000. How can we learn about developmental processes from cross-sectional studies, or can we? Am. J. Psychiatry 157 (2), 163–171. http://dx.doi.org/10.1176/appi.ajp.157.2.163.

Larsen, B., Luna, B., 2018. Adolescence as a neurobiological critical period for the development of higher-order cognition. Neurosci. Biobehav. Rev. 94, 179–195. http://dx.doi.org/10.1016/j.neubiorev.2018.09.005.

Laube, C., Lorenz, R., van den Bos, W., 2020a. Pubertal testosterone correlates with adolescent impatience and dorsal striatal activity. Dev. Cogn. Neurosci. 42, 100749. http://dx.doi.org/10.1016/j.dcn.2019.100749.

Laube, C., van den Bos, W., Fandakova, Y., 2020b. The relationship between pubertal hormones and brain plasticity: Implications for cognitive training in adolescence. Dev. Cogn. Neurosci. 100753. http://dx.doi.org/10.1016/j.dcn.2020.100753.

Lee, M.D., 2011. How cognitive modeling can benefit from hierarchical Bayesian models. J. Math. Psych. 55 (1), 1–7. http://dx.doi.org/10.1016/j.jmp.2010.08.013.

Lee, D., Seo, H., Jung, M.W., 2012. Neural basis of reinforcement learning and decision making. Annu. Rev. Neurosci. 35, 287–308. http://dx.doi.org/10.1146/annurev-neuro-062111-150512.

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., Palminteri, S., 2017. Behavioural and neural characterization of optimistic reinforcement learning. Nat. Hum. Behav. 1 (4), 0067. http://dx.doi.org/10.1038/s41562-017-0067.

Lloyd, A., McKay, R., Sebastian, C.L., Balsters, J.H., 2020. Are adolescents more optimal decision-makers in novel environments? Examining the benefits of heightened exploration in a patch foraging paradigm. Dev. Sci. n/a (n/a), e13075. http://dx.doi.org/10.1111/desc.13075, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/desc.13075.

Lourenco, F., Casey, B., 2013. Adjusting behavior to changing environmental demands with development. Neurosci. Biobehav. Rev. 37 (9), 2233–2242. http://dx.doi.org/10.1016/j.neubiorev.2013.03.003.

Master, S.L., Eckstein, M.K., Gotlieb, N., Dahl, R., Wilbrecht, L., Collins, A.G.E., 2020. Disentangling the systems contributing to changes in learning during adolescence. Dev. Cogn. Neurosci. 41, 100732. http://dx.doi.org/10.1016/j.dcn.2019.100732.

Metha, J.A., Brian, M.L., Oberrauch, S., Barnes, S.A., Featherby, T.J., Bossaerts, P., Murawski, C., Hoyer, D., Jacobson, L.H., 2020. Separating probability and reversal learning in a novel probabilistic reversal learning task for mice. Front. Behav. Neurosci. 13, http://dx.doi.org/10.3389/fnbeh.2019.00270, Publisher: Frontiers.

Meyer, H.C., Bucci, D.J., 2016. Age differences in appetitive Pavlovian conditioning and extinction in rats. Physiol. Behav. 167, 354–362. http://dx.doi.org/10.1016/j.physbeh.2016.10.004.

Morris, L.S., Kundu, P., Dowell, N., Mechelmans, D.J., Favre, P., Irvine, M.A., Robbins, T.W., Daw, N., Bullmore, E.T., Harrison, N.A., Voon, V., 2016. Fronto-striatal organization: Defining functional and microstructural substrates of behavioural flexibility. Cortex 74, 118–133. http://dx.doi.org/10.1016/j.cortex.2015.11.004.

Mulder, J., Wagenmakers, E.-J., 2016. Editors' introduction to the special issue "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments". J. Math. Psych. 72, 1–5. http://dx.doi.org/10.1016/j.jmp.2016.01.002.

Nassar, M.R., Rumsey, K.M., Wilson, R.C., Parikh, K., Heasly, B., Gold, J.I., 2012. Rational regulation of learning dynamics by pupil-linked arousal systems. Nature Neurosci. 15 (7), 1040–1046. http://dx.doi.org/10.1038/nn.3130, Number: 7 Publisher: Nature Publishing Group.

Natterson-Horowitz, D.B., Bowers, K., 2019. Wildhood: the Astounding Connections Between Human and Animal Adolescents. Scribner, New York.

Navarro, D.J., 2019. Between the devil and the deep blue sea: Tensions between scientific judgement and statistical model selection. Comput. Brain Behav. 2 (1), 28–34. http://dx.doi.org/10.1007/s42113-018-0019-z.

Newman, L.A., McGaughy, J., 2011. Adolescent rats show cognitive rigidity in a test of attentional set shifting. Dev. Psychobiol. 53 (4), 391–401. http://dx.doi.org/10.1002/dev.20537, eprint: https://onlinelibrary.wiley.com/doi/pdf/10.100.

Niv, Y., 2009. Reinforcement learning in the brain. J. Math. Psych. 53 (3), 139–154.

Nussenbaum, K., Hartley, C.A., 2019. Reinforcement learning across development: What insights can we draw from a decade of research? Dev. Cogn. Neurosci. 40, 100733. http://dx.doi.org/10.1016/j.dcn.2019.100733.

O'Doherty, J.P., Lee, S.W., McNamee, D., 2015. The structure of reinforcement-learning mechanisms in the human brain. Curr. Opin. Behav. Sci. 1, 94–100. http://dx.doi.org/10.1016/j.cobeha.2014.10.004.

Op de Macks, Z.A., Bunge, S.A., Bell, O.N., Wilbrecht, L., Kriegsfeld, L.J., Kayser, A.S., Dahl, R.E., 2016. Risky decision-making in adolescent girls: The role of pubertal hormones and reward circuitry. Psychoneuroendocrinology 74, 77–91. http://dx.doi.org/10.1016/j.psyneuen.2016.08.013.

O'Reilly, J.X., Schüffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S., 2013. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc. Natl. Acad. Sci. USA 110 (38), E3660–3669. http://dx.doi.org/10.1073/pnas.1305373110.

Palminteri, S., Kilford, E.J., Coricelli, G., Blakemore, S.-J., 2016. The computational development of reinforcement learning during adolescence. PLoS Comput. Biol. 12 (6), http://dx.doi.org/10.1371/journal.pcbi.1004953.

Palminteri, S., Wyart, V., Koechlin, E., 2017. The importance of falsification in computational cognitive modeling. Trends Cogn. Sci. 21 (6), 425–433. http://dx.doi.org/10.1016/j.tics.2017.03.011.

Perfors, A., Tenenbaum, J.B., Griffiths, T.L., Xu, F., 2011. A tutorial introduction to Bayesian models of cognitive development. p. 61.

Petersen, A.C., Crockett, L., Richards, M., Boxer, A., 1988. A self-report measure of pubertal status: Reliability, validity, and initial norms. J. Youth Adolesc. 17 (2), 117–133. http://dx.doi.org/10.1007/BF01537962.

Peterson, D.A., Elliott, C., Song, D.D., Makeig, S., Sejnowski, T.J., Poizner, H., 2009. Probabilistic reversal learning is impaired in Parkinson's disease. Neuroscience 163 (4), 1092–1101. http://dx.doi.org/10.1016/j.neuroscience.2009.07.033.

Piekarski, D.J., Boivin, J.R., Wilbrecht, L., 2017a. Ovarian hormones organize the maturation of inhibitory neurotransmission in the frontal cortex at puberty onset in female mice. Curr. Biol.: CB 27 (12), 1735–1745.e3. http://dx.doi.org/10.1016/j.cub.2017.05.027.

Piekarski, D.J., Johnson, C.M., Boivin, J.R., Thomas, A.W., Lin, W.C., Delevich, K., M Galarce, E., Wilbrecht, L., 2017b. Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? Brain Res. 1654 (Pt B), 123–144. http://dx.doi.org/10.1016/j.brainres.2016.08.042.

Pitt, M.A., Myung, I.J., 2002. When a good fit can be bad. Trends Cogn. Sci. 6 (10), 421–425. http://dx.doi.org/10.1016/S1364-6613(02)01964-2.

RCoreTeam, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Romer, D., Hennessy, M., 2007. A biosocial-affect model of adolescent sensation seeking: The role of affect evaluation and peer-group influence in adolescent drug use. Prev. Sci. 8 (2), 89. http://dx.doi.org/10.1007/s11121-007-0064-7.

Rosenbaum, G., Grassie, H., Hartley, C., 2020. Valence biases in reinforcement learning shift across adolescence and modulate subsequent memory. http://dx.doi.org/10.31234/osf.io/n3vsr, type: article.

Salvatier, J., Wiecki, T.V., Fonnesbeck, C., 2016. Probabilistic programming in python using PyMC3. PeerJ Comput. Sci. 2, e55. http://dx.doi.org/10.7717/peerj-cs.55.

Sarkka, S., 2013. Bayesian Filtering and Smoothing. Cambridge University Press, Cambridge, http://dx.doi.org/10.1017/CBO9781139344203.

Schlagenhauf, F., Huys, Q.J., Deserno, L., Rapp, M.A., Beck, A., Heinze, H.-J., Dolan, R., Heinz, A., 2014. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. Neuroimage 89 (100), 171–180. http://dx.doi.org/10.1016/j.neuroimage.2013.11.034.

Sercombe, H., 2014. Risk, adaptation and the functional teenage brain. Brain Cogn. 89, 61–69. http://dx.doi.org/10.1016/j.bandc.2014.01.001.

Shepard, R., Beckett, E., Coutellier, L., 2017. Assessment of the acquisition of executive function during the transition from adolescence to adulthood in male and female mice. Dev. Cogn. Neurosci. 28, 29–40. http://dx.doi.org/10.1016/j.dcn.2017.10.009.

Simon, N.W., Gregory, T.A., Wood, J., Moghaddam, B., 2013. Differences in response initiation and behavioral flexibility between adolescent and adult rats. Behav. Neurosci. 127 (1), 23–32. http://dx.doi.org/10.1037/a0031328.

Simonsohn, U., 2018. Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. Adv. Methods Pract. Psychol. Sci. 1 (4), 538–555. http://dx.doi.org/10.1177/2515245918805755, Publisher: SAGE Publications Inc.

Solway, A., Botvinick, M., 2012. Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. Psychol. Rev. 119 (1), 120–154. http://dx.doi.org/10.1037/a0026435.

Somerville, L.H., Casey, B., 2010. Developmental neurobiology of cognitive control and motivational systems. Curr. Opin. Neurobiol. 20 (2), 236–241. http://dx.doi.org/10.1016/j.conb.2010.01.006.

Somerville, L.H., Sasse, S.F., Garrad, M.C., Drysdale, A.T., Abi Akar, N., Insel, C., Wilson, R.C., 2017. Charting the expansion of strategic exploratory behavior during adolescence. J. Exp. Psychol. [Gen.] 146 (2), 155–164. http://dx.doi.org/10.1037/xge0000250, Place: US Publisher: American Psychological Association.

Minto de Sousa, N., Gil, M.S.C.d.A., McIlvane, W.J., 2015. Discrimination and reversal learning by toddlers aged 15-23 months. Psychol. Rec. 65 (1), 41–47. http://dx.doi.org/10.1007/s40732-014-0084-1.

Sowell, E.R., Peterson, B.S., Thompson, P.M., Welcome, S.E., Henkenius, A.L., Toga, A.W., 2003. Mapping cortical change across the human life span. Nature Neurosci. 6 (3), 309–315. http://dx.doi.org/10.1038/nn1008.

Steinberg, L., 2005. Cognitive and affective development in adolescence. Trends Cogn. Sci. 9 (2), 69–74. http://dx.doi.org/10.1016/j.tics.2004.12.005.

Sugawara, M., Katahira, K., 2021. Dissociation between asymmetric value updating and perseverance in human reinforcement learning. Sci. Rep. 11 (1), 3574. http://dx.doi.org/10.1038/s41598-020-80593-7, Number: 1 Publisher: Nature Publishing Group.

Sutton, R.S., Barto, A.G., 2017. Reinforcement Learning: an Introduction, second ed. MIT Press, Cambridge, MA; London, England.

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

Toga, A.W., Thompson, P.M., Sowell, E.R., 2006. Mapping brain maturation. Trends Neurosci. 29 (3), 148–159. http://dx.doi.org/10.1016/j.tins.2006.01.007.

Uttal, W.R., 1990. On some two-way barriers between models and mechanisms. Percept. Psychophys. 48 (2), 188–203. http://dx.doi.org/10.3758/BF03207086.

van den Bos, W., Bruckner, R., Nassar, M.R., Mata, R., Eppinger, B., 2017. Computational neuroscience across the lifespan: Promises and pitfalls. Dev. Cogn. Neurosci. http://dx.doi.org/10.1016/j.dcn.2017.09.008.

van den Bos, W., Cohen, M.X., Kahnt, T., Crone, E.A., 2012. Striatum–medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. Cerebral Cortex 22 (6), 1247–1255. http://dx.doi.org/10.1093/cercor/bhr198.

van den Bos, W., Hertwig, R., 2017. Adolescents display distinctive tolerance to ambiguity and to uncertainty during risky decision making. Sci. Rep. 7 (1), 40962. http://dx.doi.org/10.1038/srep40962, Number: 1 Publisher: Nature Publishing Group.

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

Watanabe, S., 2013. A widely applicable Bayesian information criterion. J. Mach. Learn. Res. 14 (Mar), 867–897.

Webb, B., 2001. Can robots make good models of biological behaviour? Behav. Brain Sci. 24 (6), 1033–1050. http://dx.doi.org/10.1017/S0140525X01000127, Publisher: Cambridge University Press.

Wilson, R.C., Collins, A.G., 2019. Ten simple rules for the computational modeling of behavioral data. ELife 8, e49547. http://dx.doi.org/10.7554/eLife.49547, (T. E. Behrens, Ed.) Publisher: eLife Sciences Publications, Ltd.

Xia, L., Master, S.L., Eckstein, M.K., Baribault, B., Dahl, R.E., Wilbrecht, L., Collins, A.G.E., 2021. Modeling changes in probabilistic reinforcement learning during adolescence. PLoS Comput. Biol. 17 (7), e1008524. http://dx.doi.org/10.1371/journal.pcbi.1008524, Publisher: Public Library of Science.

Xia, L., Master, S., Eckstein, M., Wilbrecht, L., Collins, A.G.E., 2020. Modeling changes in probabilistic reinforcement learning during adolescence. 17, (7), e1008524. http://dx.doi.org/10.1371/journal.pcbi.1008524, Publisher: Public Library of Science,

Yaple, Z.A., Yu, R., 2019. Fractionating adaptive learning: A meta-analysis of the reversal learning paradigm. Neurosci. Biobehav. Rev. 102, 85–94. http://dx.doi.org/10.1016/j.neubiorev.2019.04.006.

Yu, A.J., Dayan, P., 2005. Uncertainty, neuromodulation, and attention. Neuron 46 (4), 681–692. http://dx.doi.org/10.1016/j.neuron.2005.04.026.