RESEARCH ARTICLE

Modeling changes in probabilistic reinforcement learning during adolescence

Liyu Xia[1], **Sarah L. Master**[2], **Maria K. Eckstein**[3], **Beth Baribault**[3], **Ronald E. Dahl**[4], **Linda Wilbrecht**[3,5], **Anne Gabrielle Eva Collins**[3,5]*

1 Department of Mathematics, University of California Berkeley, Berkeley, California, United States of America, **2** Department of Psychology, New York University, New York, New York, United States of America, **3** Department of Psychology, University of California Berkeley, Berkeley, California, United States of America, **4** School of Public Health, University of California Berkeley, Berkeley, California, United States of America, **5** Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, California, United States of America

* annecollins@berkeley.edu

Citation: Xia L, Master SL, Eckstein MK, Baribault B, Dahl RE, Wilbrecht L, et al. (2021) Modeling changes in probabilistic reinforcement learning during adolescence. PLoS Comput Biol 17(7): e1008524. https://doi.org/10.1371/journal.pcbi.1008524

Editor: Stefano Palminteri, Ecole Normale Superieure, FRANCE

Received: November 3, 2020

Accepted: May 26, 2021

Published: July 1, 2021

Data Availability Statement: All data is available on an Open Science Framework repository: https://osf.io/wq4te/.

Funding: This work was funded by National Science Foundation SL-CN grant #1640885 to RD, AGEC, and LW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

In the real world, many relationships between events are uncertain and probabilistic. Uncertainty is also likely to be a more common feature of daily experience for youth because they have less experience to draw from than adults. Some studies suggest probabilistic learning may be inefficient in youths compared to adults, while others suggest it may be more efficient in youths in mid adolescence. Here we used a probabilistic reinforcement learning task to test how youth age 8-17 (N = 187) and adults age 18-30 (N = 110) learn about stable probabilistic contingencies. Performance increased with age through early-twenties, then stabilized. Using hierarchical Bayesian methods to fit computational reinforcement learning models, we show that all participants' performance was better explained by models in which negative outcomes had minimal to no impact on learning. The performance increase over age was driven by 1) an increase in learning rate (i.e. decrease in integration time scale); 2) a decrease in noisy/exploratory choices. In mid-adolescence age 13-15, salivary testosterone and learning rate were positively related. We discuss our findings in the context of other studies and hypotheses about adolescent brain development.

Author summary

Adolescence is a time of great uncertainty. It is also a critical time for brain development, learning, and decision making in social and educational domains. There are currently contradictory findings about learning in adolescence. We sought to better isolate how learning from stable probabilistic contingencies changes during adolescence with a task that previously showed interesting results in adolescents. We collected a relatively large sample size (297 participants) across a wide age range (8–30), to trace the adolescent developmental trajectory of learning under stable but uncertain conditions. We found that age in our sample was positively associated with higher learning rates and lower choice exploration. Within narrow age bins, we found that higher saliva testosterone levels were associated with higher learning rates in participants age 13–15 years. These findings can help us better isolate the trajectory of maturation of core learning and decision making processes during adolescence.

Introduction

In the everyday world, perfectly predictable outcomes are rare. Yet, we need to track important events and their relationships to other events and actions. For example, we might want to learn where the best place to obtain food is, or where a potential mate likes to hang out—this might help us decide where to go, expecting a positive outcome to occur frequently, but not always. Our ability to learn about these probabilistic relationships is crucial for our daily life and decision making.

This challenge needs to be met by the developing brain, especially during adolescence [4–9]. Naively, one might assume that the brain simply gets better at this (and possibly all) forms of learning with brain maturation. However, what does *better* mean in this context? Most learning mechanisms are subject to tradeoffs between speed and stability. Fast learning may be suitable for a highly certain environment with deterministic relationships/statistics, but can lead to impulsive behavior in a more uncertain environment with probabilistic relationships/statistics [2, 10]. By contrast, slower learning that integrates over a longer time scale may lead to more robust and stable performance in probabilistic environments. During development, there may be periods where one form of learning is emphasized over the other. Changes could be gradual and monotonic, or show sharp steps when driven by factors such as hormonal changes at puberty onset [11]. There may also be inverted U shapes [3, 7, 12], that peak to support a sensitive period when specific information is available in the environment and/or when an organism needs to accomplish its transition to independence [13–15].

To study how learning changes across adolescence, we used the theoretical framework of reinforcement learning (RL). Computational RL models assume that we estimate the long-term values of an action in a given state by integrating over time the feedback we receive for choosing this action in this state, through a trial-and-error process [16]. RL has greatly enhanced our understanding of human behavior and the neural processes that underlie learning and decision-making in both certain and uncertain environments [17–21]. Moreover, RL processes offer a quantitative parameterization of individual differences: for example, RL decision noise may capture exploratory choice [4]; RL learning rates control the time scale of integration of rewards, with potential asymmetries between positive and negative outcomes [22, 23]; and RL forgetting parameters may capture memory dependent processes [12].

For these reasons, RL has been previously used to probe developmental changes in learning and decision making, including during adolescence [2–5, 12, 24, 25]. While there has been some consensus on certain developmental trends, such as lower decision noise with age [4], in general, developmental results in both how learning behavior changes and in how RL processes (and parameters) change are highly variable and dependent on the specific tasks used [4, 26].

To study how learning under uncertainty changes during adolescence, we used the *Butterfly task* [2], where participants needed to learn probabilistic associations that were stable throughout the task. We collected data from a sample of 297 participants across a wide age range (8–30), over-sampling participants age 8–18 to focus on the adolescent period (see S1 Fig for detailed breakdown of age group by sex). In fact, in this same sample, we conducted a total of four tasks that varied across multiple dimensions (such as deterministic/probabilistic feedback, stable/volatile contingencies, memory load, etc.), with the initial motivation to address the issue introduced by task heterogeneity [4]. However, the focus of this paper is on the Butterfly task alone. While we mainly present results from the Butterfly task, we also discuss comparisons and relationships with two other tasks in this sequence of four tasks [3, 12].

The Butterfly task tests participants' ability to learn probabilistic associations between four butterflies and two possible preferred flowers from reward feedback. This task has been used in developmental studies before [2], and produced an intriguing result showing adolescent performance was greater than adults in a two group design (N = 41 adolescents age 13–17 and N = 31 adults age 20–30). We sought to further investigate performance in this task during development with a larger sample that would enable evaluation of the trajectory of development from age 8–30 and examine the role of puberty in changes in performance.

To evaluate the potential role of gonadal hormones and pubertal development in driving changes in learning, we also measured pubertal development and saliva testosterone (see Participants, S1 Text: Saliva collection and testosterone testing). We expected to observe an inverted U shape in performance that peaked in mid adolescence [27], coinciding with previously observed peaks in nucleus accumbens activation in response to rewarding outcomes [7, 28]. Previous studies have also found positive relationships between adolescent testosterone levels and nucleus accumbens activation [7, 29, 30]; therefore, we expected that pubertal development might explain the timing of any observed peak. A further motivation to conduct this study was to investigate the possibility that participants of different ages were differently sensitive to positive and negative outcomes, something that has been observed in other studies [31], but was not investigated previously using the Butterfly task [2].

Contrary to our predictions, we found no evidence for adolescent performance advantage in our version of the Butterfly task. Instead, we found performance increased through early adulthood, then stabilized. We used hierarchical Bayesian methods to fit computational RL models to the trial-by-trial data (see Hierarchical model fitting) and examined how participants integrated information across trials and made decisions. Increases in performance with age were explained by an increase in learning from rewarded outcomes and a decrease in exploration. These findings are largely consistent with studies of learning and decision making in other tasks that show steady improvement in performance across adolescent development [2, 4, 12]. We compare and contrast with findings that show adolescents outperforming adults [2, 3] to shed light on the conditions when adolescents vs. adults may show performance advantages in learning.

Materials and methods

Ethics statement

All procedures were approved by the Committee for the Protection of Human Subjects (CPHS number, community participants: 2016–06-8925; student participants: 2016–01-8280) at the University of California, Berkeley (UCB).

Participants

A total of 297 (151 female) participants completed the task: 187 children and adolescents (age 8–18) from the community, 55 UCB undergraduate students (age 18–25), and 55 adults (age 25–30) from the community. Participants under 18 years old and their guardians provided their informed assent or written permission; participants over 18 provided informed written consent themselves.

We assessed biological sex (self-reported) and pubertal development for children and adolescents through saliva samples and through self-report with the pubertal development questionnaire, from which we calculated testosterone levels (T1, see S1 Text: Saliva collection and testosterone testing) and Puberty Development Score (PDS, [32]) respectively. The correlation between pubertal measures and age was very strong as expected (see S11 Fig).

Community participants were compensated with a $25 Amazon gift card for completing the experimental session, and an additional $25 for completing optional take-home saliva samples; undergraduate participants received course credit for participation. All participants were pre-screened for the absence of present or past psychological and neurological disorders.

Experimental design

This task was the third in a sequence of four tasks that participants completed in the experimental session [12]. The task was a contextual two-armed bandit task with binary feedback: there were four stimuli (blue, purple, red, and yellow butterflies) and two bandits (pink and white flowers). Participants needed to figure out the preferred flower for each of the four butterflies through trial and error. Each butterfly had a preferred flower, which remained fixed throughout the experiment.

On each trial (Fig 1A), participants were presented one butterfly and two flowers. They needed to choose a flower within 7s using a video game controller. They were instructed to respond as quickly as possible. The chosen flower would stay on the screen for 1s. If participants correctly chose the preferred flower, they would receive positive feedback (*Win!*) with 80% chance; however, 20% of the time the other flower would be the rewarding one, resulting in negative feedback (*Lose!*). The feedback stayed on the screen for 2s. If participants received the 'Win!' feedback, they received 1 point, whereas 'Lose!' meant 0 points. Participants were instructed to obtain as many points as possible and they could always see the total number of points earned so far on the upper right corner of the screen. The total amount of points won was not translated to a real-life reward such as money. There were 30 trials for each butterfly, resulting in a total of 120 trials. The butterfly-flower mapping, position of flowers, sequence of butterflies and the probabilistic feedback were pre-randomized and counterbalanced across participants.

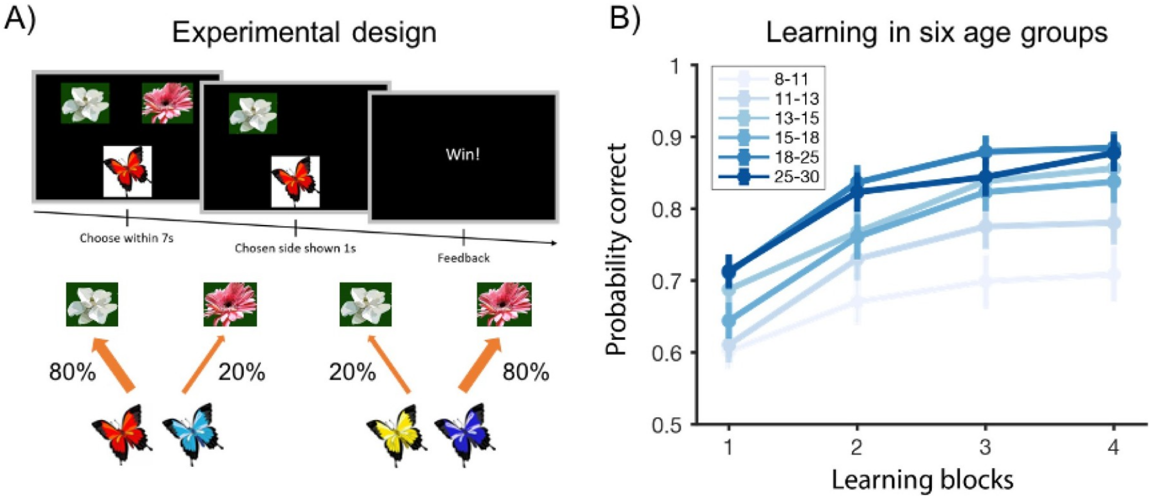


Fig 1. Experimental design and overall performance. (A) On each trial, participants needed to choose the butterfly's preferred flower. Each butterfly's preferred flower stayed the same throughout the experiment. If participants correctly picked the butterfly's preferred flower, they observed a *Win!* feedback with 80% chance, and *Lose!* otherwise. The other choice delivered positive feedback only with 20% chance. Schematics were adapted from [2]. (B) Average probability of a correct choice over four 30-trial learning blocks. Learning curves showed all age groups learned the task, and that performance generally improved with age group.

https://doi.org/10.1371/journal.pcbi.1008524.g001

Exclusion criteria

One participant under 18 was excluded because only 18 out of 120 trials were completed, resulting in 296 participants. We also excluded participants who were overall more likely to change their choice of flower for a given butterfly (switch) than repeat it after receiving positive feedback, which suggested that they either did not understand the task or were not engaged in it. 20 participants under 18 and one undergraduate participant were excluded due to this criterion. Note that all behavioral results presented later in Overall performance, Reaction time, and Mixed-effect logistic regression hold with the original sample of 296 participants.

To further conservatively identify participants who were not engaged in the task, we excluded participants who had worse than chance performance and satisfied one of the following "low data quality" criteria: (1) high proportion of trials where the participant picked the same choice as the previous trial, (2) high proportion of trials where the participant changed their choice, (3) presence of too long sequences of trials where the participant kept choosing the same flower, and (4) high proportion of missing trials. All of those criteria were determined by elbow points (see S1 Text: Exclusion criteria details), and indicated a lack of reactivity to the task's inputs. Applying these *conjunctive* criteria resulted in further exclusion of 11 participants (S1 Table), bringing the total number of *on task* participants for later analysis to 264 (138 female), with 157 participants under 18 (see S1 Table for breakdown of exclusion by age). All behavioral and modeling results presented later in Results hold with weaker exclusion criteria (i.e. with the sample of 275 participants).

Model-independent analysis

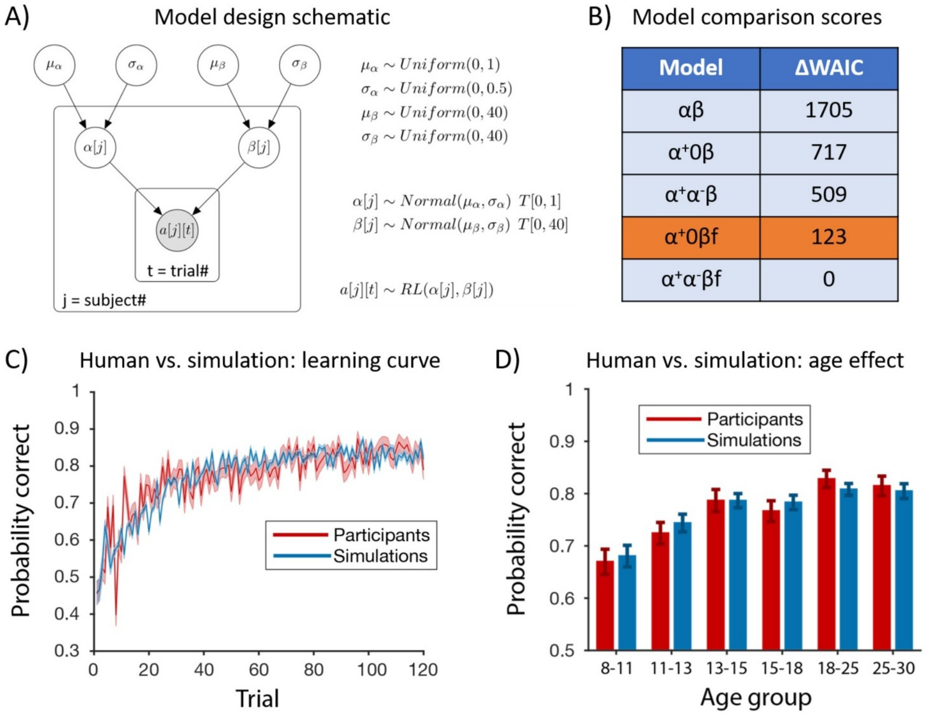
For each participant in each trial, we recorded whether they chose the butterfly's preferred flower (*correct* choice or not), and whether they received reward or not (win vs. lose), which was different due to the probabilistic nature of the task. As an aggregate measure of performance, we computed average accuracy within each of the four 30-trial learning blocks for each participant. We also computed median and standard deviation of reaction time within each learning block. We ran (linear and quadratic) regression to assess whether those behavioral metrics changed with age and pubertal measures.

We also calculated the proportion of trials (p) among all 120 trials where participants correctly chose each butterfly's preferred flower as an overall performance measure. Because this proportion was not normally distributed across participants (Kolmogorov–Smirnov test, $p = 0.003$), we instead used log odds ($\log \frac{p}{1-p}$) for all later statistical tests. The log odds were normally distributed (Kolmogorov–Smirnov test, $p = 0.26$).

We used the median reaction time for each participant as a speed measure. Because reaction time was not normally distributed across participants (Kolmogorov–Smirnov test, $p = 0.02$), for all later statistical tests, we used log-transformed reaction time, which was normally distributed (Kolmogorov–Smirnov test, $p = 0.8$).

To visualize age effects (Figs 1 and 2), we broke participants under age 18 into four equal-size groups within each sex respectively, and then combined both sexes (see S1 Text: Pubertal effects extended). The boundaries for the four age groups were approximately 8–11, 11–13, 13–15, and 15–18 (the exact boundaries for each age group and sex can be found in S2 Table). Together with two age groups above 18 (18–25 for students and 25–30 for community participants), we had a total of six age groups.

Going beyond aggregate measures across trials, we also ran a mixed effect logistic regression to predict participants' choices on a trial-by-trial basis and tested how previous reward history and delay affected learning and decision making. Specifically, for each trial, we defined the *reward history*, r, as the number of trials that participants had previously received a *Win!*



Model	ΔWAIC
$\alpha\beta$	1705
$\alpha^+0\beta$	717
$\alpha^+\alpha^-\beta$	509
$\alpha^+0\beta f$	123
$\alpha^+\alpha^-\beta f$	0

Fig 2. Hierarchical Bayesian modeling and model comparison. (A) We illustrate the hierarchical model design with the $\alpha\beta$ model as an example, We assumed that parameters of individual participants came from the same group level distribution, which is truncated normal parametrized by the group-level mean and standard deviation (μ_α and σ_α for α; μ_β and σ_β for β). The group-level mean and standard deviation followed weakly informative priors (uniform and bounded). The parameters for each individual participant are used in the likelihood of each action on every trial based on the $\alpha\beta$ model. $T[m, n]$ indicates truncation of distribution (e.g. the learning rate, α, is bounded by [0, 1]). The filled circle represented observed variable (in this case, participants' choices on each trial); unfilled circles represented latent variables (in this case, group and individual model parameters). (B) We calculated WAIC for model comparison. ΔWAIC represents the difference between the WAIC of the considered model and the model with the lowest (i.e. best) WAIC. While the $\alpha^+\alpha^-\beta f$ model had the lowest WAIC score, the α^- parameter was not recoverable and showed signs of overfitting (see S1 Text: Model comparison extended). We thus focus on the $\alpha^+0\beta f$ model. (C-D) We used fitted parameters from the $\alpha^+0\beta f$ model to generate simulated trajectories. The simulated performance captures the average learning curve throughout the experiment (C), and replicates the age effect (D).

https://doi.org/10.1371/journal.pcbi.1008524.g002

feedback for the current butterfly; we also defined *delay*, d, as the number of intervening trials since the last time the participant encountered the same butterfly and got rewarded. We then used the lme4 package in R to test $p(correct) = logit(1 + r + d + (1 + r + d|sub))$, where *sub* represented random effects of individual participants respectively. All regressors were z-scored.

We analyzed whether the random effects and other behavioral measures varied with age using linear and quadratic regressions. While quadratic regressions could provide initial evidence for nonlinear relationships, to ascertain true (inverse) U shapes, we used two-line regression [33]. Two-line regression detects a sign change of the regression slopes for low and high values of age respectively, without assuming quadratic as the true relationship between age and the dependent variables.

Using multiple linear regressions and ANOVA tests, we also assessed the effect of sex as a way to control for a potential, known source of variance and thus improve the generalizability of our findings [34]. Note that this is particularly important when assessing age and puberty effects during adolescence, as puberty has known different timelines across sexes, and different hormonal markers (also visible in our sample).

Computational models

While the logistic regression described above (Model-independent analysis) could serve as a descriptive model for participants' trial-by-trial choices, we also used computational RL modeling to obtain a more quantitative and mechanistic understanding of participants' trial-by-trial learning. We applied six variants of RL models, then used the parameter estimates of the best fitting model as the basis for inference.

Classic RL ($\alpha\beta$). Our simplest RL model was the $\alpha\beta$ model, with just two free parameters, α (learning rate) and β (inverse temperature). The $\alpha\beta$ model used Q-learning to compute $Q(b, a)$, as the expected value of choosing flower a for butterfly b. On trial t, the probability of choosing a was computed by transforming the Q-value with a softmax function:

$$P(a|b) = \frac{exp(\beta Q_t(b,a))}{\sum_{i=1}^{2} exp(\beta Q_t(b,a_i))}, \tag{1}$$

where $Q_t(b, a)$ was the Q-value until trial t. The inverse temperature parameter β thus controls how exploratory/stochastic the decision making process is, with higher β resulting in more deterministic choices. After observing reward r_t (0 for "Lose!" or 1 for "Win!"), the Q-value $Q(b, a)$ was updated through the classic delta rule:

$$Q_{t+1}(b,a) = Q_t(b,a) + \alpha RPE, \tag{2}$$

where $RPE = r_t - Q_t(b, a)$ is the reward prediction error. Note that this delta rule can also be rewritten as:

$$Q_{t+1}(b,a) = (1-\alpha)Q_t(b,a) + \alpha r_t, \tag{3}$$

which shows the updated Q-value (Q_{t+1}) as a linear combination of past estimates (Q_t) and the most recent reward (r_t). Thus, the learning rate parameter α is often interpreted as a time integration constant, controlling how much of the past estimate contributes to the current estimate. For example, $\alpha = 1$ would result in one-shot learning, i.e. set Q-value to be identical to the reward feedback each trial, resulting in an integration time scale of one trial (and no information about any other past trials). Smaller α results in integrating reward information across more trials from present into the past. Note that this time integration constant occurs only over trials in which the specific stimulus is present, rather than all trials.

We initialized all Q-values to the uninformative value of 0.5 (the average of positive and negative feedback) for this model and all other models under consideration.

RL with asymmetric learning rates ($\alpha^+\alpha^-\beta$). The $\alpha^+\alpha^-\beta$ model differed from the $\alpha\beta$ model by using two distinct learning rate parameters, α^+ and α^-. Recent literature suggests that humans learn from positive and negative feedback to different degrees, and even with potentially different neural mechanisms [23]. RL models with asymmetric learning rates have also been widely used and examined in theoretical [22, 35] and developmental [1, 24, 31, 36, 37] contexts, especially in studies with probabilistic tasks.

Having both α^+ and α^- allowed the model to have different sensitivity to positive and negative RPE [38]. Specifically, in Eq 2, α^+ was used when $RPE > 0$, and α^- otherwise.

Asymmetric RL with $\alpha^- = 0$ ($\alpha^+0\beta$). The $\alpha^+0\beta$ model was the same as the $\alpha^+\alpha^-\beta$ model, except that the α^- parameter was set to 0. This change made the model insensitive to negative feedback. We included this model because of the observation that the fitted values of the α^- parameter from the $\alpha^+\alpha^-\beta$ model were very small and not recoverable (see S1 Text: Model comparison extended).

RL with forgetting ($\alpha\beta f$). The $\alpha\beta f$ model builds upon the $\alpha\beta$ model by including an additional forgetting parameter, f. On each trial, after applying the delta learning rule Eq 2, Q-values decay toward the uninformative starting value of 0.5, implementing a forgetting process:

$$Q_{t+1}(b,a) = (1-f) * Q_{t+1}(b,a) + f * 0.5. \tag{4}$$

Eq 4 is implemented for all butterfly-flower pairs except the butterfly and the selected flower on the current trial. Note that forgetting thus occurs on every trial (in contrast to integration via learning rate which occurs only on stimulus-specific trials).

Asymmetric RL with forgetting ($\boldsymbol{\alpha^{+}\alpha^{-}\beta f}$). The $\alpha^{+}\alpha^{-}\beta f$ model has both asymmetric learning rates for positive and negative feedback and the forgetting parameter.

Asymmetric RL with $\boldsymbol{\alpha^{-} = 0}$ and forgetting ($\boldsymbol{\alpha^{+}0\beta f}$). For factorial design, we included the $\alpha^{+}0\beta f$ model, which builds upon the $\alpha^{+}\alpha^{-}\beta f$ model by setting $\alpha^{-} = 0$.

Hierarchical model fitting

We fitted all RL models using hierarchical Bayesian methods [39] jointly to all participants, instead of to each participant independently. To illustrate the hierarchical model design (Fig 2A), we use the simplest model, $\alpha\beta$, as an example. We specified weakly informative priors for the mean and standard deviation of the group-level learning rate (μ_α and σ_α) and the group-level inverse temperature (μ_β and σ_β). We assumed that these group-level parameters were all uniformly distributed over the natural ranges of the parameters (for example, we truncated μ_α at 0 and 1, since we know the learning rate parameter is between 0 and 1). We then assumed that the parameters for each participant were drawn from a prior distribution defined by the group-level parameters: for example, $\alpha[j]$ for participant j was drawn from a normal distribution $Normal(\mu_\alpha, \sigma_\alpha)$ truncated at 0 and 1. Individual participants' parameters were then used in the likelihood of each participant's actions on each trial ($a[j][t]$) according to the $\alpha\beta$ model, where j and t indicate participant number and trial number, respectively.

The hierarchical model made the likelihood intractable [40], but it can be well approximated by sampling. We used No-U-Turn sampling, a state-of-the-art Markov Chain Monte Carlo (MCMC) algorithm implemented in the probabilistic programming language Stan [41], to sample from the joint posterior distribution of model parameters for all participants. Compared to the classic participant-wise maximum likelihood estimation approach, hierarchical model fitting with MCMC provides more stable point estimates for individual participants and allows natural inference of effects on parameters at the group level [42].

For each model, we ran 4 MCMC chains, with each chain generating 4000 samples (after 1000 warmup samples), resulting in 16000 samples per model for later inference. We assessed convergence for all models using the *matstanlib* library [43]. In particular, we ensured that $\hat{R}$ statistics for all free parameters were below 1.05; that the effective sample sizes (ESS) for all free parameters were more than 25 × the number of chains; and that samples generally were not the result of divergent transitions. Note that these criteria are more stringent than the standard criteria for convergence of $\hat{R} \leq 1.1$ and ESS $>$ 5× number of chains as per [39]. Among all 6 models, only the $\alpha\beta f$ model was unable to converge when fitted hierarchically, thus we fitted the $\alpha\beta f$ model independently for each participant. The results presented later for the $\alpha\beta f$ model all came from non-hierarchical fitting. We also fitted the other 5 models non-hierarchically in order to compare with the $\alpha\beta f$ model, and summarized the results in the supplement (S6 Fig).

Hierarchical Bayesian modeling also provides a natural way to test for potential age effects on model parameters. Specifically, we incorporated the regression model using age to predict model parameters into the original graphical model (Fig 2A), and directly sampled regression coefficients for age jointly with other model parameters.

Same as before, we assumed that the model parameters for individual participants followed a truncated normal distribution with a group-level standard deviation, but now we replace the prior on the group-level mean with a regression statement with respect to age. For example, to probe linear effects of age on α, we assumed that the parameter $\alpha[j]$, used to compute the likelihood of participant j's choices, followed $Normal(\alpha_{intercept} + \alpha_{linear} * age[j], \sigma_\alpha), T\,[0, 1]$, where $age[j]$ was the z-scored age of participant j, and $\alpha_{intercept}$, α_{linear} were regression coefficients for which we set weakly informative priors. To probe quadratic effects, we just further included $\alpha_{quadratic} * age[j]^2$.

To test for effects of age on the model parameters, we examined whether the posterior distribution of all 16000 samples for the linear (α_{linear}) and quadratic ($\alpha_{quadratic}$) regression coefficients were significantly different from 0.

Parameter and model identifiability and validation

We verified that model parameters and models themselves were identifiable using generate and recover procedures (see S1 Text: Model comparison extended, [44]). We validated models by simulating models with fitted parameters 100 times per participants, and comparing model simulations with behavior (S4 Fig).

Results

Overall performance

To assess learning progress and potential age effects, we first calculated the proportion of correct trials within each of the four 30-trial learning blocks for 6 age groups (Fig 1B). As indicated in S1 Text: Pubertal effects extended, we grouped all participants under 18 into four equal-sized bins (N = 39, 39, 39, 40). The other two groups were undergraduate participants (age 18–25, N = 53) and adult community participants (age 25–30, N = 54).

All age groups exhibited learning over the course of the experiment. Specifically, we found a significant main effect of age group and block on participants' performance (two-way mixed-effects ANOVA, age group: $F(5, 255) = 8.5$, $p < 0.0001$; block: $F(3, 765) = 136$, $p < 0.0001$). There was no interaction between age group and block (two-way mixed-effects ANOVA: $F(15, 765) = 1$, $p = 0.49$). This shows that participant's performance improved as the experiment progressed, and older participants generally outperformed younger participants.

To further characterize the effect of age on overall performance, we computed the proportion of correct trials over all 120 trials. We found that the overall performance of 13–18 year-olds (top two quartiles) was significantly higher than of 8–13 year-olds (bottom two quartiles; unpaired t-test, $t(1, 155) = 3.5$, $p = 0.0001$), and significantly lower than of 18–25 year-olds (unpaired t-test, $t(1, 130) = 2.5$, $p = 0.01$). However, there was no significant difference (unpaired t-test, $t(1, 105) = 0.2$, $p = 0.8$) between the performance of 25–30 year-olds and 18–25 year-olds (Figs 1B and 2D).

To examine the continuous relationship between participants' performance and age, we ran a regression analysis using age to predict performance (Fig 3A). We found that including a quadratic term improved model fit (sequential ANOVA: $F(1, 261) = 7.6$, $p = 0.006$). The regression analysis revealed linear and quadratic effects of age on performance (linear: $\beta_{age} = 0.05$, 95% CI = [0.03, 0.07]; quadratic: $\beta_{age^2} = -0.004$, 95% CI = [−0.007, −0.001]). There was no effect of sex or its interaction with age (multiple linear regression, both p's > 0.45).

To identify whether the quadratic effect was indicative of an inverse U shape, we conducted the two-line regression [33]. We found the break point at around 19 years old, before which $\beta_{age} = 0.11$, $z = 5.24$, $p < 0.0001$, and after which $\beta_{age} = 0$, $z = 0.14$, $p = 0.9$. This indicates that performance linearly increased with age through adolescence and stabilized in early

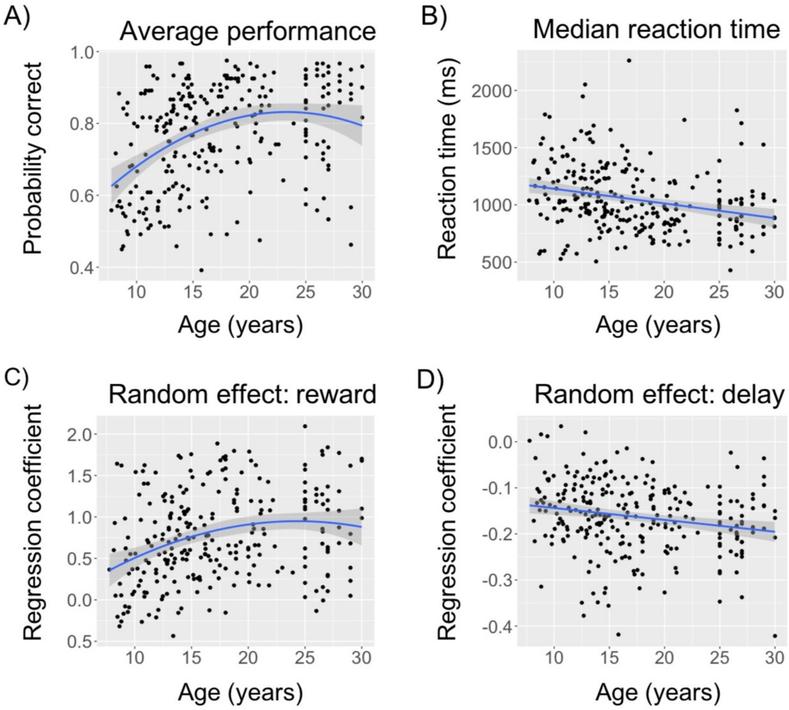


Fig 3. Age effects on participants' behavior. Scatter plot of age (x-axis) and (A) probability of choosing the correct response, (B) median reaction time, (C) random effect for reward history, and (D) random effect of delay. Each black dot represents one participant. The blue curve represents linear/quadratic regression line. There was no effect of sex in any analysis. Shaded region represents 95% confidence interval.

https://doi.org/10.1371/journal.pcbi.1008524.g003

adulthood, showing that the quadratic effect reflected a linear increase then stable process rather than an inverse U-shape.

Reaction time

We also computed the median (Fig 3B) and standard deviation of reaction time for each participant. We found a linear effect of age on median reaction time ($\beta_{age} = -0.01$, 95% CI = [−0.02, −0.006]). This suggests that participants reacted faster with age, confirming previous results [12]. Adding a quadratic term did not improve model fit (sequential ANOVA: $F(1, 261) = 1$, $p = 0.3$). We also found a linear effect of age on the standard deviation of reaction time (linear regression: $\beta_{age} = -0.02$, 95% CI = [−0.03, −0.01]); adding a quadratic term improved model fit (sequential ANOVA: $F(1, 261) = 17$, $p < 0.0001$; quadratic regression: $\beta_{age^2} = 0.003$, 95% CI = [0.002, 0.005]). Two-line regression revealed a break point at 19 years old, before which we found $\beta_{age} = -0.05$, $z = -4.15$, $p < 0.0001$, and after which we found $\beta_{age} = 0.03$, $z = 1.88$, $p = 0.06$. This indicates that the variability in reaction time decreased with age, and this decrease stabilizes in early adulthood and might even invert, consistent with previous findings [12, 45]. There was no significant effect of sex on the median reaction time (unpaired t-test, median: $t(1, 262) = 0.4$, $p = 0.7$), but female participants had a significantly smaller standard deviation than male participants (unpaired t-test: $t(1, 262) = 2.72$, $p = 0.0085$).

These results indicate better performance and faster responses in older participants, ruling out speed-accuracy tradeoffs. Both age group (Fig 1B) and continuous age (Fig 3A) analyses revealed a nonlinear saturating relationship between age and performance.

Mixed-effect logistic regression

To better probe trial-by-trial learning dynamics, we used reward history and delay to predict the probability of a correct choice on each trial in a mixed-effect logistic regression. We found significant fixed effects of reward history and delay ($\beta_r = 0.8$, $\beta_d = -0.17$, both p's < 0.0001). This suggests that participants were more likely to pick the preferred flower as they received more reward feedback for the butterfly (reinforcement learning effect), and encountered the butterfly more recently (forgetting effect).

We found linear and quadratic effects (linear: $\beta_{age} = 0.02$, 95% CI = [0.01, 0.03]; quadratic: $\beta_{age^2} = -0.002$, 95% CI = [−0.004, −0.0003]; sequential ANOVA: $F(1, 261) = 5$, $p < 0.02$) of age on the random effect of reward history (Fig 3C). Two-line regression reveals a break point at around 21 years old, before which we found $\beta_{age} = 0.04$, $z = 3.84$, $p = 0.0001$, and after which we found $\beta_{age} = 0.02$, $z = 0.75$, $p = 0.45$. Therefore, similar to the trend we observed for overall performance (Fig 3A), participants' sensitivity to reward increased with age and stabilized in early adulthood. We also found that participants became more sensitive to delay with age, shown by the linear effect ($\beta_{age} = -0.003$, 95% CI = [−0.004, −0.001]) of age on the random effect of delay (Fig 3D). Adding a quadratic term did not improve model fit (sequential ANOVA, $F(1, 261) = 3$, $p = 0.08$).

Computational modeling

We used computational modeling and model comparison to obtain a mechanistic understanding of participants' trial-by-trial learning and decision making. We fitted all participants jointly using hierarchical Bayesian modeling [39] combined with sampling [41] for approximating the likelihood function (see Hierarchical model fitting).

Model comparison. We used WAIC to compare the relative fit of models at the population level [46], an information criterion that penalizes model complexity appropriately for hierarchical Bayesian models. WAIC is fully Bayesian and invariant to reparametrization. Smaller WAIC indicates a better fit to the data, controlling for complexity. Since the $\alpha\beta f$ model was unable to converge when fitted hierarchically, we compared the other five hierarchical models using WAIC. Model comparison results for all six models fitted non-hierarchically with BIC can be found in the supplement (S6 Fig).

The $\alpha^+\alpha^-\beta f$ model with asymmetric learning rates and the forgetting parameter had the lowest (best) WAIC score (Fig 2B). However, a generate and recover procedure [44] showed that the α^- parameter values in the $\alpha^+\alpha^-\beta f$ model were very close to 0 (see S1 Text: Model comparison extended), and that they were not adequately recoverable (and therefore unsuitable to use as the basis for inference, see S2A and S3 Figs). Consequently, we focus on the model with the next best WAIC score, $\alpha^+0\beta f$, which could be successfully recovered from (see S1 Text: Model comparison extended), for further analysis. Note that conclusions for the α^+, β, and f parameters remain the same if we used the $\alpha^+\alpha^-\beta f$ model instead. Furthermore, the α^- parameters were generally very small for the fitted $\alpha^+\alpha^-\beta f$ model (S2A Fig). This suggests that participants were learning either very little from negative feedback or not at all. Moreover, the $\alpha^+0\beta f$ model resulted in better model validation (see S1 Text: Model comparison extended, S4 Fig), suggesting that $\alpha^+\alpha^-\beta f$ at the population level might be overfitting.

We validated the best-fitting model, $\alpha^+0\beta f$, by simulating synthetic choice trajectories from fitted parameters (i.e., by generating posterior predictive distributions; Fig 2C and 2D) [47]. Model simulations captured the average learning curve throughout the entire experiment (Fig 2C) and age effects on overall performance (Fig 2D).

Age differences in model parameters. With the winning model $\alpha^+0\beta f$, we next asked which computational processes drove the changes in performance over age by testing how

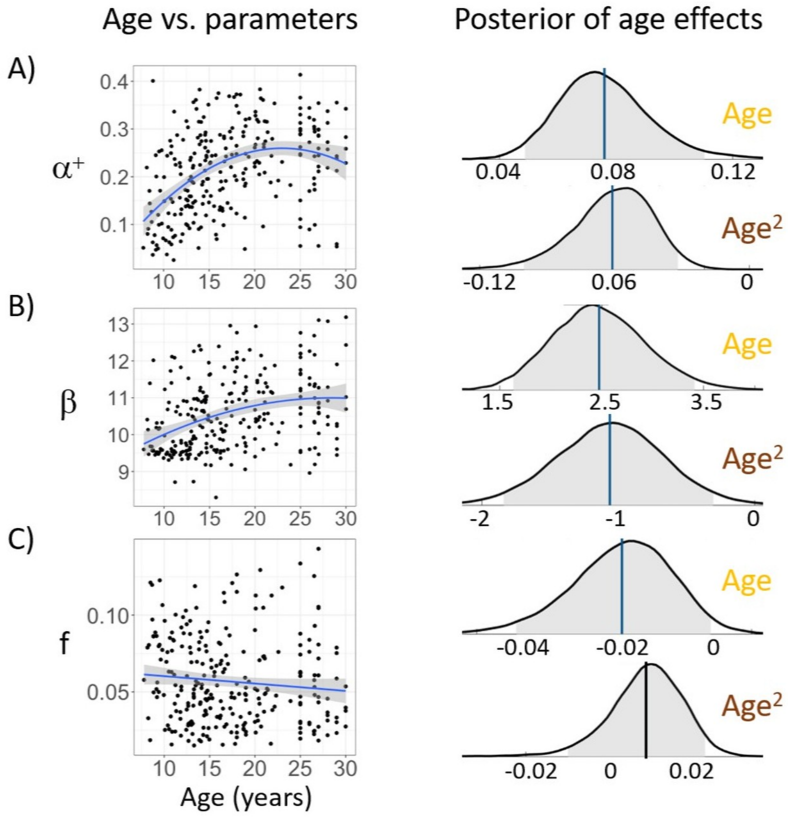


Fig 4. Age effects on model parameters. We directly incorporated age-related parameters into MCMC sampling to test within the hierarchical Bayesian modeling framework whether age had a linear or quadratic effect on all three model parameters: α^+ (A), β (B), f (C). Left panel: individual parameters from the original $\alpha^+0\beta f$ model plotted against age. For visualization, we included a regression line; the shaded region indicates 95% CI. Right: distribution of posterior samples for linear (top, yellow) and quadratic (bottom, brown) regression coefficients. The vertical line represents the mean of all samples, with blue indicating an effect being present (i.e., 95% CI not including 0), and black indicating no effect. Shaded region shows 95% confidence interval.

https://doi.org/10.1371/journal.pcbi.1008524.g004

model parameters changed with age. We adapted hierarchical Bayesian modeling to probe effects of age on model parameters. Specifically, we incorporated the regression of age as a predictor of model parameters into the hierarchical Bayesian model (Fig 2A), and directly sampled regression coefficients for age jointly with other model parameters (see Hierarchical model fitting).

To test for effects of age on the model parameters, we examined whether the 95% credible interval (CI) of the posterior samples for each of the linear and quadratic regression coefficients did or did not include 0, where 0 indicates no effect (Fig 4). We found linear and quadratic effects of age on α^+ (linear coefficient 95% CI = [0.05, 0.11]; quadratic coefficient 95% CI = [−0.1, −0.03]) and β (linear CI = [1.6, 3.4]; quadratic CI = [−1.9, −0.3]). The trajectory of quadratic change over age for α^+ and β closely mimicked that for overall performance (Fig 3A). We also found marginally linear (Fig 3C), but not quadratic effects of age on f (linear CI = [−0.04, 0.001], $p = 0.066$; quadratic CI = [−0.01, 0.02]), with the forgetting parameter potentially decreasing over age.

Pubertal effects

To study whether pubertal development also affected participants' learning and decision making, we used pubertal measures (pubertal development score PDS and testosterone level T1) to predict behavioral measures and fitted model parameters (Fig 5). Note that these analyses were only conducted on participants under 18.

Regarding the behavioral measures (Fig 5A, 5B, 5E and 5F), we found that T1, but not PDS, had a marginal linear effect on overall performance (linear regression. PDS: $\beta_{PDS} = 0.1$, 95% CI

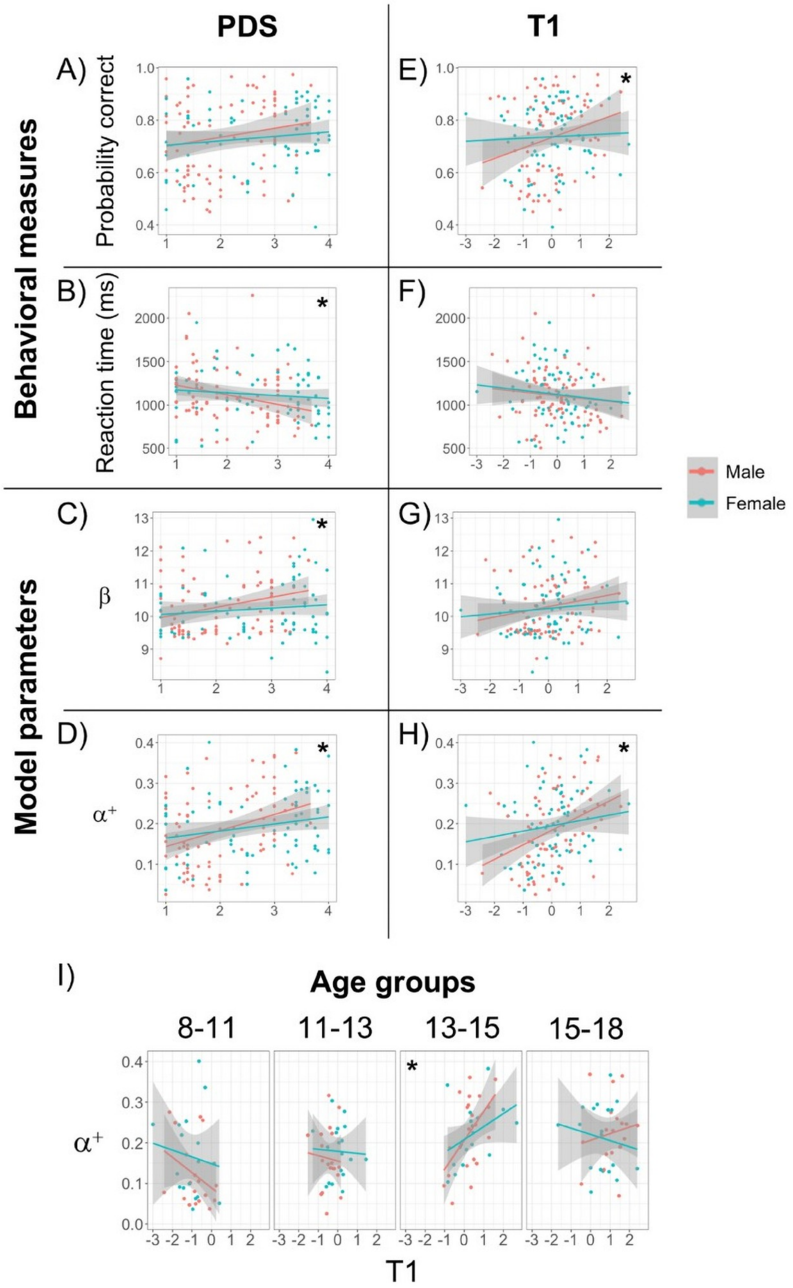


Fig 5. Pubertal effects on behavioral measures and model parameters. Scatter plots of (A) Puberty development scale (PDS) vs. overall performance, (B) PDS vs. median reaction time, (C) PDS vs. α^+, and (D) PDS vs. β. (E-H) Same as (A-D) but with z-scored testosterone level (T1). (I) Scatter plot of T1 vs. α^+ in each of the four age groups younger than age 18. Each colored dot represents one participant under 18. Color indicates self-reported sex. There were no significant effects of sex or interactions. The lines represent best fitting linear regression lines. Shaded region represents 95% confidence interval. * indicates a significant main effect of PDS or T1 (see main text).

https://doi.org/10.1371/journal.pcbi.1008524.g005

= [−0.04, 0.24], $p = 0.18$; T1: $\beta_{T1} = 0.14$, 95% CI = [0, 0.27], $p = 0.05$). PDS, but not T1, had a marginal linear effect on reaction time (linear regression. PDS: $\beta_{PDS} = -0.05$, 95% CI = [−0.1, 0], $p = 0.05$; T1: $\beta_{T1} = -0.03$, 95% CI = [−0.07, 0.02], $p = 0.25$). When controlling for age in the regression, the effect of T1 on performance disappeared (multiple linear regression, $p(T1) = 0.53$, $p(age) = 0.001$). However, for the effect of PDS on reaction time, adding age did not improve model fit (sequential ANOVA, $F(1, 153) = 0.1$, $p = 0.76$). Note that these results are generally consistent with our previous findings relating age and behavioral measures (Fig 3), as expected given the strong correlation between age and puberty measures (see S11 Fig), albeit noisier. There was no effect of sex in any of the regressions (multiple linear regression, all p's > 0.16).

For the fitted model parameters (Fig 5C, 5D, 5G and 5H), we found that both PDS and T1 had a linear effect on α^+ (linear regression. PDS: $\beta_{PDS} = 0.02$, 95% CI = [0.01, 0.04], $p = 0.0007$; T1: $\beta_{T1} = 0.02$, 95% CI = [0.01, 0.04], $p = 0.0006$), and no effect on the forgetting parameter f (linear regression; both p's > 0.5). There was a linear effect of PDS, but not T1, on β (linear regression. PDS: $\beta_{PDS} = 0.15$, 95% CI = [0.003, 0.29], $p = 0.046$; T1: $\beta_{T1} = 0.13$, 95% CI = [−0.01, 0.27], $p = 0.08$). However, the effects of PDS and T1 on α^+ and β disappeared when adding age into the regression (multiple linear regression, all p's > 0.55), while age remained the only significant predictor (multiple linear regression, all p's < 0.027). There was no effect of sex in any of the regressions (multiple linear regression, all p's > 0.2).

To further explore the effect of PDS and T1 on model parameters while controlling for age, we performed the same regression using PDS or T1 to predict model parameters within each of the four age groups under 18 (Fig 5I and S12 Fig), within which age and puberty measures were less correlated (S4 Table). We found that within the third age group (Fig 5I) (age 13–15), there was a positive effect of T1 on α^+ (linear regression: $\beta_{T1} = 0.05$, 95% CI = [0.01, 0.08], $p = 0.005$). This remained significant when correcting for multiple comparisons (two parameters by four groups). This T1 effect remained when controlling for age in the regression (multiple linear regression: $\beta_{T1} = 0.05$, $p(T1) = 0.006$, $p(age) = 0.95$), as T1 and age were decorrelated in this age group ($\rho = -0.03$, $p = 0.85$; S4 Table). T1 did not provide additional explanatory power for β and f parameters (see S1 Text: Pubertal effects extended).

Discussion

How do humans learn to make choices when the outcome is uncertain? To learn probabilistic contingencies, humans need to integrate information over multiple trials to avoid overreacting to noise in the environment. But to learn efficiently, they also need to pay attention to recent information. Here, we investigated how humans trade off these constraints across development, what the underlying computational mechanisms that support such learning are, and how they change during adolescence.

At the population level, computational model comparison (Fig 2B) suggested that two mechanisms modulated learning of probabilistic contingencies. First, participants did not treat positive and negative feedback identically; rather, they had a strong bias to learn more from positive, and little to none from negative feedback. This asymmetry has been widely observed in previous studies [12, 24, 31], potentially due to differential mechanisms integrating positive and negative feedback [48]. Second, we found that learning was better explained by including a forgetting mechanism: more intervening trials between two iterations of a choice decreased the strength of past information [12].

Consistent with the age effects observed in previous work using tasks with probabilistic [3] and deterministic [12] feedback, our behavioral and modeling results suggest that learning in a stable probabilistic task environment changed markedly from childhood to adulthood. In particular, we found that overall performance increased with age, stabilising in early adulthood. This behavioral pattern was mirrored by the learning rate parameter (α^+) as well as inverse temperature (β), a parameter indicating a decrease in noise or exploration in choice.

Our observations that learning rate α^+ and inverse temperature β increase with age and pubertal development during adolescence are generally consistent with previous work using the deterministic learning task *RLWM*, tested in the same participants as shown here [12], and a probabilistic task with same the same overall task structure as the Butterfly task, but different feedback methods [2]. However, we did not find higher performance in adolescents than adults, as had been observed in this previous Butterfly task study [2] (Fig 3A). Even when using the same age bins as [2], which limited our number of participants to N = 84 13–18-year-old adolescents and N = 86 20–30-year-old adults, we instead found that the performance in 20–30-year-olds was significantly higher than 13–18-year-olds (unpaired t-test, $t(168) = 2.3$, $p = 0.02$). Because our 18–25 year-olds were differently recruited than the rest of our sample, we additionally compared the 13–18-year-old adolescents to N = 54 25–30-year-old adults (i.e. not including the undergraduate participants between 20–25 years old). We still found significantly better performance in adults (unpaired t-test, $t(136) = 2.2$, $p = 0.03$).

The finding in [2] was interpreted as "an upside" to slower learning that led to more robust integration over time of information, and thus higher overall performance under uncertainty at younger ages. Indeed, lower learning rates can be more optimal in probabilistic tasks than higher learning rates. However, the relationship between learning rates and performance when learning probabilistic contingencies is complex and non-monotonic: it follows an inverse U-shape, as very low learning rates lead to integrating information too slowly, but high learning rates lead to being too susceptible to noisy feedback (see S1 Text: Nonlinear relationship between performance and model parameters, [44]). Furthermore, the inverse U-shape itself is dependent on the degree of exploration and forgetting ([2, 4, 44], see supplementary simulations in S7 and S8 Figs). Learning rates were smaller in our study compared to [2]: the group level mean for α^+ in our sample was 0.18, whereas in [2], the mean was around 0.3 and 0.6 for adolescents and adults respectively (Fig 2B in [2]). In higher ranges of learning rates [2], an increase in learning rate could result in a decrease in performance (right side of the inverse U-shape), while in our lower range, it could lead to an increase in performance (left side of the inverse U-shape). Thus, the two studies are consistent in identifying an increase in learning rate with age, but over a different range of learning rate values (0.3 vs. 0.6), leading to opposite effects on performance. Indeed, in our study, the parameter trajectory with age corresponded to a slow improvement towards more "optimal" behavior, as defined by correct performance in the task (S7 Fig).

Moreover, we modeled learning from positive and negative feedback asymmetrically [4], as opposed to the symmetric learning rate in [2]. In particular, our winning model $\alpha^+0\beta f$ did not learn from negative feedback at all. A high α^- can also result in worse asymptotic performance in the Butterfly task (see S9 Fig), resulting in more switching from the preferred flower. Note that when using the same model as in [2] with symmetric learning rate, i.e. the $\alpha\beta$ model, we found similar age effects (see S1 Text: Age effects in the $\alpha\beta$ model) on the α and β parameters as on the α^+ and β parameters in the winning model $\alpha^+0\beta f$.

Therefore, while we found a similar trend as in [2] that learning rates increased with age (Fig 4A), our learning rate values were much smaller, and the resulting trend in overall performance was different. Note that this difference in the range of learning rates could be a result of differences in the task specifics (our experiment did not have a memory retrieval aspect with novel images or brain imaging; our task was also the third in a sequence of four tasks). Differences in performance could also stem from differences in socioeconomic status (SES) and education level between the groups recruited to each study. For example, our 18–25 year-olds were undergraduate students, who may have a different education level than the 25–30 year-old community participants in our study or the adults sampled in [2].

The incentivization for undergraduate participants (course credit) and community participants (monetary) was also different. Furthermore, participants' performance (total points earned) was not translated into real-life reward such as money. While there have been studies showing that the human brain treats primary and secondary reinforcers similarly [49–52], the importance of incentivization remains a controversial topic in decision making research. It is thus unclear whether there are developmental differences regarding how the incentive structure in the Butterfly task might motivate participants differently at different ages.

Another potential limitation of our sample is that the majority of the excluded participants were under 18, which could contain meaningful variance. This might be due to the fact that younger participants were a bit more likely to be distracted during the task/had a more difficult time understanding the task logistics, although most (about 90%) of our younger participants were able to understand and engage well in the task to pass this criterion. In particular, we excluded participants who, for a given butterfly, were more likely to switch than stay after positive feedback (see S1 Text: Exclusion criteria details), because these participants showed no reward sensitivity and were thus likely "off-task". Since we focused on developmental changes in learning, we found it necessary and helpful for later modeling analysis to exclude participants by this criterion. The parameters from those participants would not be interpretable since the RL models all assume trial-by-trial learning from reward feedback, which these participants' behavior demonstrated a lack of. Moreover, even if we included them in modeling, they would likely have very low learning rates (since they are not sensitive to positive feedback) and low β (random/noisy behavior), which could only strengthen our findings. Finally, since the study is cross-sectional rather than longitudinal, age could also be confounded by birth year.

Nevertheless, our results support other previous developmental findings. In particular, we also found a decrease in exploration with age [1, 12], and an increase in learning rate previously observed in both deterministic [12] and probabilistic learning tasks [3]. Note that other studies have observed a decrease in learning rates (e.g., single-learning-rate models: [25, 53, 54]; models with asymmetric learning rates: [24, 31, 55]) or no change [36, 37]. These differences are potentially due to different task structures, samples, and modeling choices. For a more comprehensive review, see [4]).

While we found that performance increased during adolescence and stabilized in early adulthood in this stable probabilistic learning task, a probabilistic switching task in the same sample of participants [3] found a pronounced inverse U shape in overall performance, which peaked at age 13–15. We conclude that this difference in age of peak performance in these two tasks stems from the reliance or lack of reliance on negative outcomes. The stable associations in the Butterfly task might encourage the participants to focus mostly on positive feedback (although this is not optimal based on the simulations in S9 Fig, as α^- in the low range can improve performance), whereas in a volatile task setting [3], negative feedback was crucial for identifying when the correct action switched. This suggests that even with the same sample of participants in two probabilistic tasks, task stability / volatility greatly changed participants' behavioral strategies. For this sample of participants, the volatile condition in [3] gave the 13–15 year old adolescents an edge over adults, while the stable condition in the Butterfly Task gave young adults an edge over adolescents.

While we found that random effects of delay on performance (calculated from the mixed effect logistic regression, which is descriptive) became more pronounced with age (Fig 3D), computational model fitting in contrast showed that forgetting parameters became weaker (Fig 4C). One possible interpretation for this apparent contradiction might relate to two simultaneous changes. First, adults might rely more on working memory processes [19] for probabilistic tasks [56], which manifested in a strengthened effect of delay. However, the decay of these memory processes might also decrease with age [12], which could be captured here by the decrease in the forgetting parameter. Thus, younger participants might show a weaker effect of delay not because their memory system was forgetting less (it was forgetting more), but because they used their working memory system less in this task [12, 57], and instead relied more on slower but more robust learning systems.

While we found that pubertal measures did not explain much additional variance compared to age in model parameters (see S1 Text: Pubertal effects extended), we found that testosterone level T1 had a significant positive effect on α^{+} within the third age group of 13–15 years (Fig 5I). Several explanations for this time limited observation are possible: a) gonadal hormone effects are stronger at this time of mid puberty, b) the other individual drivers of variation are more consistent at this time allowing detection of puberty related effects, or c) this result is a type I error. We favor hypothesis a) and b) because this observation about learning rate is broadly consistent with several studies [7, 29, 30] which report a positive relationship between testosterone levels and nucleus accumbens bold activity in response to rewards in mid adolescence. These data combined suggest a putative link between testosterone, nucleus accumbens activity, and learning rate in mid adolescence. Our testosterone related findings may also be relevant to a putative sensitive period for social learning driven by gonadal hormones in adolescence [13, 58]. Future experiments may test if these relationships between learning rate and testosterone in this non-social context are replicated and/or magnified in a social context.

Overall, work on the role of puberty and learning is currently in an early phase of understanding. It is likely that there are gonadal hormone dependent and independent aspects of development in the brain that will need to be disentangled [59]. A longitudinal design will have stronger statistical power to isolate puberty dependent effects [11]. Other sources of hormones and neuropeptides may also contribute to coordinate developmental change across the body and cumulative experience may also contribute. Age is less noisy to measure than pubertal development or hormones, but age is not a satisfactory explanation at the proximate level of analysis which aims to identify upstream biological mechanisms.

Conclusion

In conclusion, we sought to examine the development of learning in a stable probabilistic environment using a large adolescent and young adult sample with continuous age in the 8–30 range. Combining behavioral analysis and computational modeling, we showed developmental gains in performance through early adulthood that were explained by an increase in learning from rewarded outcomes (corresponding to a narrower time scale of information integration) and a decrease in exploration. These data and models help explain why learning and decision making differ during development and why a 'one-size-fits-all' approach may not equally serve youth at different stages.

Supporting information

S1 Text. Supplementary results.
(PDF)

S1 Fig. Demographics.
(TIF)

S2 Fig. Generate and recover.
(TIF)

S3 Fig. Successful generate and recover of $\boldsymbol{\alpha^-}$ in a higher and healthier range.
(TIF)

S4 Fig. Model validation.
(TIF)

S5 Fig. Generate and recover for the age regression coefficients in hierarchical modeling.
(TIF)

S6 Fig. Flat model comparison of all six models per age group. We calculated the difference between the BIC for each of the six models with $\alpha^+0\beta f$ model per participant, represented by ΔBIC on the y-axis. Color represents 6 age groups. Results show that the $\alpha^+0\beta f$ model is the winning model in all groups consistently.
(TIF)

S7 Fig. Heat map for simulated performance of the $\boldsymbol{\alpha^+0\beta f}$ model. Overall simulated performance changes with respect to α^+ (y-axis) and β (x-axis), where each subplot corresponds to $f = 0 - 0.2$ from left to right. Black rectangle highlights the local maximum within each column of each subplot (i.e. fixed β value), whereas the red rectangle highlights the global maximum.
(TIF)

S8 Fig. Heat map for simulated performance of the $\boldsymbol{\alpha^+0\beta f}$ model. Overall simulated performance changes with respect to α^+ (y-axis) and f (x-axis), where each subplot corresponds to $\beta = 5 - 15$ from left to right. Black rectangle highlights the local maximum within each column of each subplot (i.e. fixed f value), whereas the red rectangle highlights the global maximum.
(TIF)

S9 Fig. Simulated performance of the $\boldsymbol{\alpha^+\alpha^-\beta}$ model. Overall simulated performance (y-axis) changes with respect to α^- (x-axis), where each subplot corresponds to a combination of (α^+, β) values. The vertical bar corresponds to the α^+ value. The error bars show standard error across 100 simulations.
(TIF)

S10 Fig. Age effects on the $\boldsymbol{\alpha\beta}$ model parameters. We directly incorporated age-related parameters into MCMC sampling to test within the hierarchical Bayesian modeling framework whether age had a linear or quadratic effect on the fitted parameters from the $\alpha\beta$ model. We found positive linear effect of age on α and β, and negative quadratic effect of age on β. The model with quadratic age effect on α failed to converge.
(TIF)

S11 Fig. Scatter plots of pubertal measures (PDS and T1) and age.
(TIF)

S12 Fig. Pubertal effects on behavior and fitted $\boldsymbol{\alpha^+}$ parameters in each of the four age groups younger than age 18. (A) PDS vs. median reaction time. (B) T1 vs. overall performance. (C) PDS vs. α^+. (D) T1 vs. α^+.
(TIF)

S1 Table. Number of participants excluded due to each exclusion criterion for each of the four age groups.
(TIF)

S2 Table. Age boundaries for each of the four age groups under 18.
(TIF)

S3 Table. Model identifiability analysis. The rows indicate the model where the dataset was generated from, whereas the columns indicate the model used for recovery. Each entry indicates protected exceedance probability.
(TIF)

S4 Table. Statistics of pubertal measures under 18. Within each of the age group under 18, we calculated the variance of pubertal measures (PDS and T1) and their correlations to age.
(TIF)

Acknowledgments

We thank the many people who contributed to this project: Amy Zou, Lance Kriegsfeld, Celia Ford, Jennifer Pfeifer, Megan Johnson, Vy Pham, Rachel Arsenault, Josephine Christon, Shoshana Edelman, Lucy Eletel, Neta Gotlieb, Haley Keglovits, Julie Liu, Justin Morillo, Nithya Rajakumar, Nick Spence, Tanya Smith, Benjamin Tang, Talia Welte, and Lucy Whitmore. We would also like to thank our participants and their families.

Author Contributions

Conceptualization: Maria K. Eckstein, Ronald E. Dahl, Linda Wilbrecht, Anne Gabrielle Eva Collins.

Data curation: Sarah L. Master, Maria K. Eckstein, Anne Gabrielle Eva Collins.

Formal analysis: Liyu Xia.

Funding acquisition: Ronald E. Dahl, Linda Wilbrecht, Anne Gabrielle Eva Collins.

Investigation: Liyu Xia, Sarah L. Master, Anne Gabrielle Eva Collins.

Methodology: Liyu Xia, Sarah L. Master, Maria K. Eckstein, Beth Baribault, Anne Gabrielle Eva Collins.

Project administration: Linda Wilbrecht, Anne Gabrielle Eva Collins.

Resources: Anne Gabrielle Eva Collins.

Software: Sarah L. Master, Maria K. Eckstein, Beth Baribault, Anne Gabrielle Eva Collins.

Supervision: Linda Wilbrecht, Anne Gabrielle Eva Collins.

Validation: Liyu Xia, Anne Gabrielle Eva Collins.

Visualization: Liyu Xia, Anne Gabrielle Eva Collins.

Writing – original draft: Liyu Xia, Linda Wilbrecht, Anne Gabrielle Eva Collins.

Writing – review & editing: Liyu Xia, Sarah L. Master, Maria K. Eckstein, Beth Baribault, Ronald E. Dahl, Linda Wilbrecht, Anne Gabrielle Eva Collins.

References

1. Christakou A, Gershman SJ, Niv Y, Simmons A, Brammer M, Rubia K. Neural and psychological maturation of decision-making in adolescence and young adulthood. Journal of cognitive neuroscience. 2013; 25(11):1807–1823. https://doi.org/10.1162/jocn_a_00447 PMID: 23859647

2. Davidow J, Foerde K, Galván A, Shohamy D. An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. Neuron. 2016; 92(1):93–99. https://doi.org/10.1016/j.neuron.2016.08.031 PMID: 27710793

3. Eckstein MK, Master SL, Dahl RE, Wilbrecht L, Collins AGE. Understanding the Unique Advantage of Adolescents in Stochastic, Volatile Environments: Combining Reinforcement Learning and Bayesian Inference. bioRxiv. 2020.

4. Nussenbaum K, Hartley CA. Reinforcement learning across development: What insights can we draw from a decade of research? Developmental Cognitive Neuroscience. 2019; 40:100733. https://doi.org/10.1016/j.dcn.2019.100733 PMID: 31770715

5. DePasque S, Galván A. Frontostriatal development and probabilistic reinforcement learning during adolescence. Neurobiology of Learning and Memory. 2017; 143:1–7. https://doi.org/10.1016/j.nlm.2017.04.009 PMID: 28450078

6. Steinberg L. A social neuroscience perspective on adolescent risk-taking. Developmental review. 2008; 28(1):78–106. https://doi.org/10.1016/j.dr.2007.08.002 PMID: 18509515

7. Braams BR, van Duijvenvoorde AC, Peper JS, Crone EA. Longitudinal changes in adolescent risk-taking: a comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. Journal of Neuroscience. 2015; 35(18):7226–7238. https://doi.org/10.1523/JNEUROSCI.4764-14.2015 PMID: 25948271

8. Somerville LH, Jones RM, Casey B. A time of change: behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. Brain and cognition. 2010; 72(1):124–133. https://doi.org/10.1016/j.bandc.2009.07.003 PMID: 19695759

9. Walker DM, Bell MR, Flores C, Gulley JM, Willing J, Paul MJ. Adolescence and reward: making sense of neural and behavioral changes amid the chaos. Journal of Neuroscience. 2017; 37(45):10855–10866. https://doi.org/10.1523/JNEUROSCI.1834-17.2017 PMID: 29118215

10. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. Nature Neuroscience. 2007; 10(9):1214–1221. https://doi.org/10.1038/nn1954 PMID: 17676057

11. Kraemer HC, Yesavage JA, Taylor JL, Kupfer D. How can we learn about developmental processes from cross-sectional studies, or can we? American Journal of Psychiatry. 2000; 157(2):163–171. https://doi.org/10.1176/appi.ajp.157.2.163 PMID: 10671382

12. Master SL, Eckstein MK, Gotlieb N, Dahl R, Wilbrecht L, Collins AGE. Distentangling the systems contributing to changes in learning during adolescence. Developmental Cognitive Neuroscience. 2020; 41:100732. https://doi.org/10.1016/j.dcn.2019.100732 PMID: 31826837

13. Dahl RE, Allen NB, Wilbrecht L, Suleiman AB. Importance of investing in adolescence from a developmental science perspective. Nature. 2018; 554(7693):441–450. https://doi.org/10.1038/nature25770 PMID: 29469094

14. Piekarski DJ, Johnson CM, Boivin JR, Thomas AW, Lin WC, Delevich K, et al. Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? Brain research. 2017; 1654:123–144. https://doi.org/10.1016/j.brainres.2016.08.042 PMID: 27590721

15. Frankenhuis WE, Walasek N. Modeling the evolution of sensitive periods. Developmental cognitive neuroscience. 2020; 41:100715. https://doi.org/10.1016/j.dcn.2019.100715 PMID: 31999568

16. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. MIT Press; 2018.

17. Niv Y. Reinforcement learning in the brain. Journal of Mathematical Psychology. 2009; 53(3):139–154. https://doi.org/10.1016/j.jmp.2008.12.005

18. Gläscher J, Daw N, Dayan P, O'Doherty JP. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron. 2010; 66(4):585–595. https://doi.org/10.1016/j.neuron.2010.04.016 PMID: 20510862

19. Collins AG, Frank MJ. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. European Journal of Neuroscience. 2012; 35(7):1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x

20. Leong YC, Radulescu A, Daniel R, DeWoskin V, Niv Y. Dynamic interaction between reinforcement learning and attention in multidimensional environments. Neuron. 2017; 93(2):451–463. https://doi.org/10.1016/j.neuron.2016.12.040 PMID: 28103483

21. Farashahi S, Rowe K, Aslami Z, Lee D, Soltani A. Feature-based learning improves adaptability without compromising precision. Nature communications. 2017; 8(1):1768. https://doi.org/10.1038/s41467-017-01874-w PMID: 29170381

22. Cazé RD, van der Meer MA. Adaptive properties of differential learning rates for positive and negative outcomes. Biological cybernetics. 2013; 107(6):711–719. https://doi.org/10.1007/s00422-013-0571-5 PMID: 24085507

23. Lefebvre G, Lebreton M, Meyniel F, Bourgeois-Gironde S, Palminteri S. Behavioural and neural characterization of optimistic reinforcement learning. Nature Human Behaviour. 2017; 1(4):1–9. https://doi.org/10.1038/s41562-017-0067

24. Hauser TU, Iannaccone R, Walitza S, Brandeis D, Brem S. Cognitive flexibility in adolescence: Neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. NeuroImage. 2015; 104:347–354. https://doi.org/10.1016/j.neuroimage.2014.09.018 PMID: 25234119

25. Palminteri S, Kilford EJ, Coricelli G, Blakemore SJ. The computational development of reinforcement learning during adolescence. PLoS computational biology. 2016; 12(6):e1004953. https://doi.org/10.1371/journal.pcbi.1004953 PMID: 27322574

26. Van Den Bos W, Güroğlu B, Van Den Bulk BG, Rombouts SA, Crone EA. Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. Frontiers in human neuroscience. 2009; 3:52. https://doi.org/10.3389/neuro.09.052.2009 PMID: 20140268

27. van der Schaaf ME, Warmerdam E, Crone EA, Cools R. Distinct linear and non-linear trajectories of reward and punishment reversal learning during development: relevance for dopamine's role in adolescent decision making. Developmental cognitive neuroscience. 2011; 1(4):578–590. https://doi.org/10.1016/j.dcn.2011.06.007 PMID: 22436570

28. Galvan A, Hare TA, Parra CE, Penn J, Voss H, Glover G, et al. Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. Journal of Neuroscience. 2006; 26(25):6885–6892. https://doi.org/10.1523/JNEUROSCI.1062-06.2006 PMID: 16793895

29. de Macks ZAO, Moor BG, Overgaauw S, Güroğlu B, Dahl RE, Crone EA. Testosterone levels correspond with increased ventral striatum activation in response to monetary rewards in adolescents. Developmental Cognitive Neuroscience. 2011; 1(4):506–516. https://doi.org/10.1016/j.dcn.2011.06.003

30. Spielberg JM, Olino TM, Forbes EE, Dahl RE. Exciting fear in adolescence: does pubertal development alter threat processing? Developmental cognitive neuroscience. 2014; 8:86–95. https://doi.org/10.1016/j.dcn.2014.01.004 PMID: 24548554

31. van den Bos W, Cohen MX, Kahnt T, Crone EA. Striatum–Medial Prefrontal Cortex Connectivity Predicts Developmental Changes in Reinforcement Learning. Cerebral Cortex. 2012; 22(6):1247–1255. https://doi.org/10.1093/cercor/bhr198 PMID: 21817091

32. Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: Reliability, validity, and initial norms. Journal of Youth and Adolescence. 1988; 17(2):117–133. https://doi.org/10.1007/BF01537962 PMID: 24277579

33. Simonsohn U. Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. Advances in Methods and Practices in Psychological Science. 2018; 1(4):538–555. https://doi.org/10.1177/2515245918805755

34. Yarkoni T. The generalizability crisis. Preprint] PsyArXiv https://doi.org/1031234/osf.io/jqw35. 2019.

35. Katahira K. The statistical structures of reinforcement learning with asymmetric value updates. Journal of Mathematical Psychology. 2018; 87:31–45. https://doi.org/10.1016/j.jmp.2018.09.002

36. Jones RM, Somerville LH, Li J, Ruberry EJ, Powers A, Mehta N, et al. Adolescent-specific patterns of behavior and neural activity during social reinforcement learning. Cognitive, Affective, & Behavioral Neuroscience. 2014; 14(2):683–697. https://doi.org/10.3758/s13415-014-0257-z PMID: 24550063

37. Moutoussis M, Bullmore ET, Goodyer IM, Fonagy P, Jones PB, Dolan RJ, et al. Change, stability, and instability in the Pavlovian guidance of behaviour from adolescence to young adulthood. PLoS computational biology. 2018; 14(12):e1006679. https://doi.org/10.1371/journal.pcbi.1006679 PMID: 30596638

38. Frank MJ, Seeberger LC, O'Reilly RC. By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism. Science. 2004; 306(5703):1940–1943. https://doi.org/10.1126/science.1102941 PMID: 15528409

39. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Chapman and Hall/CRC; 2013.

40. Daw ND, et al. Trial-by-trial data analysis using computational models. Decision making, affect, and learning: Attention and performance XXIII. 2011; 23(1). https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

41. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. Journal of Statistical Software. 2017; 76(1):1–32. https://doi.org/10.18637/jss.v076.i01

42. Katahira K. How hierarchical models improve point estimates of model parameters at the individual level. Journal of Mathematical Psychology. 2016; 73. https://doi.org/10.1016/j.jmp.2016.03.007

43. Baribault B. matstanlib: A library of helper functions for Stan/MATLABStan; 2019.

44. Wilson RC, Collins AG. Ten simple rules for the computational modeling of behavioral data. Elife. 2019; 8:e49547. https://doi.org/10.7554/eLife.49547 PMID: 31769410

45. Larsen B, Luna B. Adolescence as a neurobiological critical period for the development of higher-order cognition. Neuroscience & Biobehavioral Reviews. 2018; 94:179–195. https://doi.org/10.1016/j.neubiorev.2018.09.005 PMID: 30201220

46. Watanabe S. A widely applicable Bayesian information criterion. Journal of Machine Learning Research. 2013; 14(Mar):867–897.

47. Palminteri S, Wyart V, Koechlin E. The Importance of Falsification in Computational Cognitive Modeling. Trends in Cognitive Sciences. 2017; 21(6):425–433. https://doi.org/10.1016/j.tics.2017.03.011 PMID: 28476348

48. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proceedings of the National Academy of Sciences. 2007; 104(41):16311–16316. https://doi.org/10.1073/pnas.0706111104 PMID: 17913879

49. Valentin VV, O'Doherty JP. Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. Journal of neurophysiology. 2009; 102(6):3384–3391. https://doi.org/10.1152/jn.91195.2008 PMID: 19793875

50. Chib VS, Rangel A, Shimojo S, O'Doherty JP. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. Journal of Neuroscience. 2009; 29 (39):12315–12320. https://doi.org/10.1523/JNEUROSCI.2575-09.2009 PMID: 19793990

51. Kim H, Shimojo S, O'Doherty JP. Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. Cerebral cortex. 2011; 21(4):769–776. https://doi.org/10.1093/cercor/bhq145 PMID: 20732900

52. McNamee D, Rangel A, O'doherty JP. Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. Nature neuroscience. 2013; 16(4):479–485. https://doi.org/10.1038/nn.3337 PMID: 23416449

53. Decker JH, Lourenco FS, Doll BB, Hartley CA. Experiential reward learning outweighs instruction prior to adulthood. Cognitive, Affective, & Behavioral Neuroscience. 2015; 15(2):310–320. https://doi.org/10.3758/s13415-014-0332-5

54. Javadi AH, Schmidt DH, Smolka MN. Adolescents adapt more slowly than adults to varying reward contingencies. Journal of cognitive neuroscience. 2014; 26(12):2670–2681. https://doi.org/10.1162/jocn_a_00677 PMID: 24960048

55. Buritica JMR, Heekeren HR, van den Bos W. The computational basis of following advice in adolescents. Journal of experimental child psychology. 2019; 180:39–54. https://doi.org/10.1016/j.jecp.2018.11.019

56. McDougle SD, Collins AG. Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning. Psychonomic bulletin & review. 2020; p. 1–20.

57. Do KT, Sharp PB, Telzer EH. Modernizing conceptions of valuation and cognitive-control deployment in adolescent risk taking. Current Directions in Psychological Science. 2020; 29(1):102–109. https://doi.org/10.1177/0963721419887361 PMID: 33758473

58. Fuhrmann D, Knoll LJ, Blakemore SJ. Adolescence as a sensitive period of brain development. Trends in cognitive sciences. 2015; 19(10):558–566. https://doi.org/10.1016/j.tics.2015.07.008 PMID: 26419496

59. Delevich K, Thomas AW, Wilbrecht L. Adolescence and "late blooming" synapses of the prefrontal cortex. In: Cold Spring Harbor symposia on quantitative biology. vol. 83. Cold Spring Harbor Laboratory Press; 2018. p. 37–43.