Learning under uncertainty changes during adolescence

Liyu Xia
jimmyxia@berkeley.edu
Department of Mathematics
Berkeley, CA 94720 USA

Sarah L. Master
sarah.master@tuebingen.mpg.de
Max Planck Institute for Biological Cybernetics
Tübingen, 72076 Germany

Maria Eckstein
maria.eckstein@berkeley.edu
Department of Psychology
Berkeley, CA 94720 USA

Linda Wilbrecht
wilbrecht@berkeley.edu
Department of Psychology
Helen Wills Neuroscience Institute
Berkeley, CA 94720 USA

Anne G. E. Collins
annecollins@berkeley.edu
Department of Psychology
Helen Wills Neuroscience Institute
Berkeley, CA 94720 USA

Abstract

As we transition from child to adult, we navigate the world differently. In this world, many of the relationships between events are unclear or uncertain because they are probabilistic in nature. We wanted to know how learning about probabilistic relationships changes with development and to interrogate the underlying processes. We investigated these questions in a probabilistic reinforcement learning task (The Butterfly Task) with 302 participants aged 8-30. We found performance in this task increased with age through early-twenties, then stabilized. Using hierarchical Bayesian methods to fit computational reinforcement learning models, we showed that this performance increase was driven by 1) an increase in learning rate (i.e. decrease in integration time horizon); 2) a decrease in exploratory choices. By contrast, forgetting rates did not change with age. We discuss our findings in the context of other studies and hypotheses about adolescent brain development.

Keywords: reinforcement learning; computational modeling; uncertainty; development

Introduction

In the everyday world, perfectly predictable outcomes are rare. Yet, we still 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 therefore crucial for our daily life and decision making. It follows that this challenge needs to be met by the developing brain. From a naive perspective, 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 more uncertain environment with probabilistic relationships/statistics (Behrens, Woolrich, Walton, & Rushworth, 2007). By contrast, slower and more integrated learning may lead to more robust and stable performance in probabilistic environments. During development, there may be periods where one form of learning may be emphasized over the other. Changes could be gradual and monotonic, but there may also be non-monotonic changes (e.g. inverted U shapes (Master et al., 2020)) that accommodate the expected increase in uncertainty in the environment with the transition to independence during adolescence (Dahl, Allen, Wilbrecht, & Suleiman, 2018).

Here, we investigate these changes across adolescence using a theoretical framework commonly used to investigate learning from reward outcomes, reinforcement learning (RL). Computational RL models assume that we estimate the long term values of actions by aggregating the feedback we receive for them over time, through a trial-and-error process (Sutton & Barto, 2018). 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 (Niv, 2009; Gläscher, Daw, Dayan, & O'Doherty, 2010; Collins & Frank, 2012).

We examined how 302 participants age 8-30 learned probabilistic relationships in the Butterfly task. The Butterfly task has been used in developmental studies before (Davidow, Foerde, Galván, & Shohamy, 2016), and tests participants' ability to learn about four butterflies preferences for two possible flowers. Each butterfly is programmed to choose one flower 80% of the time and the other 20% of the time. The challenge for the participant is to correctly predict the flower the butterfly will choose. We examined trial-by-trial learning about the preferences of the butterflies. We found that performance increased through early twenties, then stabilized, peaking at around 24 years. We next used hierarchical Bayesian methods to fit computational RL models to the trial-by-trial data (see Computational modeling) and examined how subjects 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 data are largely consistent with a general picture emerging from studies of learning and decision making across development (Davidow et al., 2016; Master et al., 2020; Nussenbaum & Hartley, 2019). We also discuss some notable differences (Davidow et al., 2016).

Methods

Participants

All procedures were approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley (UCB). A total of 302 participants completed the task: 187 children and adolescents (age 8-17) from the community, 60 UCB undergraduates (age 18-25), and 55 adults (age 25-30) from the community.

Community subjects were compensated with a $25 Amazon gift card for completing the experimental session; undergraduate participants received course credits for participation. All participants were pre-screened for the absence of present or past psychological and neurological disorders.

Experimental design

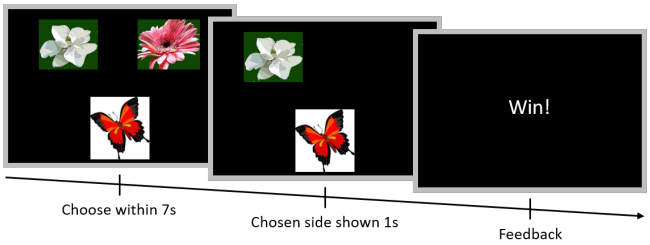

Figure 1: Experimental design. On each trial, participants needed to select the flower that the butterfly preferred. Each butterfly had the same preferred flower throughout the experiment. If participants selected the butterfly's preferred flower, they observed a *Win!* feedback with probability 0.8, and *Lose!* otherwise. For non-preferred choices, they received a *Win!* feedback with probability 0.2, and *Lose!* otherwise.

This task was one of four tasks participants completed in the experimental session (Master et al., 2020). The task involved learning from probabilistic binary feedback in a contextual 2-armed bandit task. Participants were instructed to figure out, for each of four butterflies (blue, purple, red, and yellow), which of two flowers (pink or white) the butterfly preferred through trial and error. Each butterfly had the same preferred flower throughout the experiment.

On each trial (Fig. 1), participants saw one butterfly and chose one of the two flowers using a video game controller. Participants had 7s to respond following the onset of the butterfly and the flowers, but were instructed to respond as quickly as possible. Once a flower was chosen, it stayed on the screen for 1s. If participants correctly chose the preferred flower, they were rewarded (*Win!*) 80% of the time and received negative feedback *Lose!* 20% of the time. If the other flower was chosen, participants only received reward 20% of the time. After the participants made their selection, the feedback stayed on the screen for 2s. There were 120 trials (30 trials for each butterfly) in total. The butterfly-flower mapping, position of flowers, sequence of butterflies and the probabilistic feedback were pre-randomized and counterbalanced across participants.

Exclusion criteria

To ensure that participants understood the task and stayed engaged, we excluded any participants who were more likely to change their flower choice for a given butterfly after a win trial than after a lose trial. This criterion allowed us to include every participant who showed evidence of being sensitive to feedback, even if their overall performance was close to chance. We excluded 20 community participants and 1 undergraduate participant based on this criterion. One more community participant was excluded because only 18 out of 120 learning trials were completed. Data from 5 more undergraduate participants were excluded for being outside our age range. In total, we excluded 21 participants under 18, and 6 participants in the 18-25 age range. Our final analysis included a total of 275 participants (166 under age 18).

Computational models

We used computational modeling to characterize participants' trial-by-trial responses. We used hierarchical Bayesian modeling to fit parameters and compare five models.

Classic RL ($\alpha\beta$) The $\alpha\beta$ model is the simplest Q-learning model with just 2 free parameters, α (learning rate) and β (inverse temperature), that learns to estimate $Q(b,a)$, the expected value of choosing flower a for butterfly b. All Q-values are initialized to the uninformative value of 0.5. On trial t, the probability of choosing a is computed by transforming the Q-value with a softmax:

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

where β is the inverse temperature, and $Q_t(b,a)$ is the Q-value until trial t. After reward r_t (0 for "Lose!" or 1 for "Win!") is presented, the Q-value corresponding to butterfly b and flower a is updated through the classic delta rule:

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

where α is the learning rate parameter, and $RPE = r_t - Q_t(b,a)$ is the reward prediction error.

RL with asymmetric learning rates ($\alpha^+\alpha^-\beta$) The $\alpha^+\alpha^-\beta$ model differs from the $\alpha\beta$ model by using two distinct learning rate parameters, α^+ and α^- to capture different sensitivity to wins and losses (Frank, Seeberger, & O'Reilly, 2004). Specifically, the update in equation (2) occurs with α^+ when $RPE > 0$, and α^- otherwise.

Asymmetric RL with $\alpha^- = 0$ ($\alpha^+0\beta$) The $\alpha^+0\beta$ model stems from our observation that the fitted α^- parameters from the $\alpha^+\alpha^-\beta$ model were very low, suggesting that participants might not be integrating much information from negative feedback. To test this possibility, we also included a model where we fixed $\alpha^- = 0$.

RL with forgetting ($\alpha^+0\beta f$) The $\alpha^+0\beta f$ model builds upon the $\alpha^+0\beta$ model by introducing the forgetting parameter, f. On each trial, after the learning update in equation

(2), Q-values decay toward the uninformative value of 0.5, mimicking forgetting:

$$Q_{t+1}(b,a) = (1-f) * Q_{t+1}(b,a) + f * 0.5. \quad (3)$$

Forgetting occurs for all butterfly-flower pairs except the butterfly and the selected flower on the current trial.

RL with asymmetric learning rates and forgetting ($\alpha^+\alpha^-\beta f$) The $\alpha^+\alpha^-\beta f$ model is the same as the $\alpha^+ 0\beta f$ model, except the learning rate for negative RPE, α^-, is a free parameter, and not fixed to 0.

Results

Human behavior

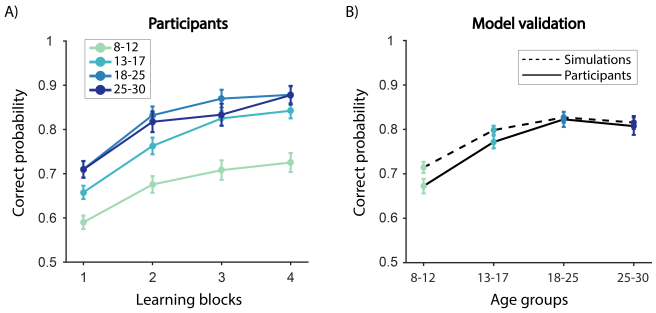

Figure 2: Performance by age group. (A) Average probability of a correct choice over 30-trial learning blocks. Learning curves show that 13-17 year-olds perform better than 8-12 year-olds, but worse than adults. (B) Overall performance of participants (solid line) and model simulations (dashed line) for each age group. The winning $\alpha^+\alpha^-\beta f$ model was used for model simulations. Colors indicate different age groups; error bars indicate standard error.

We characterized performance in terms of correct choices: trials in which the participant selected the butterfly's experimenter-defined preferred flower (different from trials in which they were rewarded). We first analyzed the average number of correct choices within each of the four 30-trial learning blocks to assess participants' learning performance. To visualize the effect of age on learning curves, we averaged participants' performance within different age groups (Fig. 2A). In particular, we grouped all participants under 18 into an age 8-12 group (N = 80) and age 13-17 group (N = 86). Undergraduate participants (age 18-25, N = 54) and adult community participants (age 25-30, N = 55) constituted the other 2 age groups.

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 (2-way mixed-effects ANOVA, age group: $F(3,264) = 15.6, p < 0.0001$; block: $F(3,792) = 133, p < 0.0001$). There was no interaction between age group and block (2-way mixed-effects ANOVA: $F(9,792) = 1.1, p = 0.33$).

To further characterize the effect of age on overall performance, we computed the proportion of correct trials over all four learning blocks (Fig. 2B). Because this proportion was not normally distributed across participants (Kolmogorov–Smirnov test, $p = 0.002$), we instead used log odds for all later statistical tests. The distribution of log odds was normally distributed (Kolmogorov–Smirnov test, $p = 0.13$).

We found that the overall performance of the 13-17 year-olds was significantly higher than 8-12 year-olds (unpaired t-test, $t(1,164) = 4.3, p < 0.0001$), and significantly lower than 18-25 year-olds (unpaired t-test, $t(1,138) = 2.5, p = 0.013$). However, there was no significant difference between the performance of 25-30 year-olds and 18-25 year-olds (unpaired t-test, $t(1,107) = 0.24, p = 0.8$).

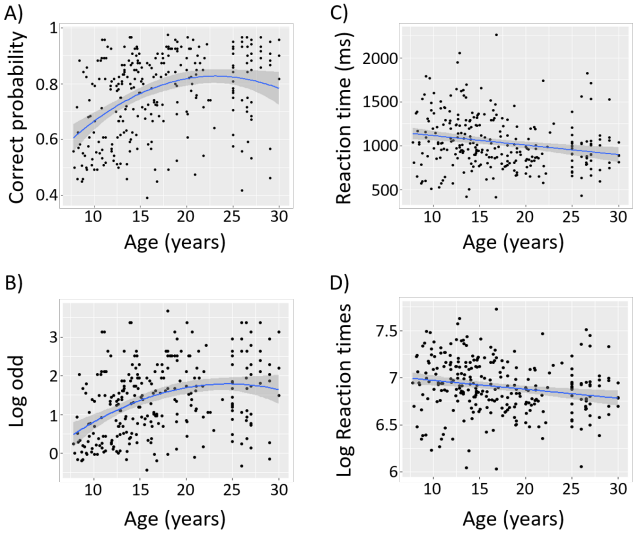

Figure 3: Age effects on participants' behavior. Scatter plot of age (x-axis) and (A) probability of choosing the correct response, (B) log odd of probability of choosing the correct response, (C) median reaction time (in milliseconds), and (D) log of median reaction time. Each black dot represents one participant. The blue curve represents the regression line (quadratic for age, linear for reaction time, see results). Shaded region represents 95% confidence interval.

To examine the continuous relationship between participants' behavior and age, we ran a multiple regression analysis (Fig. 3AB). We found that including a quadratic term improved fit in terms of the Akaike Information Criterion (AIC; AIC(linear) = 712; AIC(quadratic) = 705). The regression analysis revealed both linear and quadratic effects (linear: $\beta_{age} = 0.06, p < 0.0001$; quadratic: $\beta^2_{age} = -0.005, p = 0.002$). This indicated an inverse U-shape performance curve, with maximal performance around age 24, confirming the previous group analysis. There was no significant effect of sex or interaction with age (multiple linear regression, both p's > 0.57).

We also computed the median (Fig. 3CD) and standard

deviation of reaction time for each participant. Because reaction time was not normally distributed across participants (Kolmogorov–Smirnov test, $p = 0.02$), for all later statistical tests, we used log reaction time, which was normally distributed (Kolmogorov–Smirnov test, $p = 0.89$).

Confirming previous results (Master et al., 2020), we found a significant linear effect of age on the median of reaction time ($\beta_{age} = -0.01, p = 0.001$), indicating that reaction times became faster with age; including a quadratic term did not improve fit. We also found a significant linear effect of age on the standard deviation of reaction time (linear regression: $\beta_{age} = -0.02, p < 0.0001$); adding a quadratic term provided a better fit (AIC(linear) = 352, AIC(quadratic) = 330, $\beta^2_{age} = 0.004, p < 0.0001$). This indicates that the variability in reaction time decreased with age, and this decrease itself slowed down with age, consistent with previous findings (Master et al., 2020; Larsen & Luna, 2018). There was no significant effect of sex on the median reaction time (unpaired t-test, median: $t(1,273) = 0.77, p = 0.44$), but female participants had a significantly smaller standard deviation than male participants (unpaired t-test: $t(1,273) = 2.72, p = 0.007$).

These results indicate better performance and faster response in older participants. Age group (Fig. 2) and continuous age (Fig. 3AB) analysis both revealed an inverse-U shape with performance, suggesting that the age effect slowed down in adulthood, and might even invert.

Computational modeling

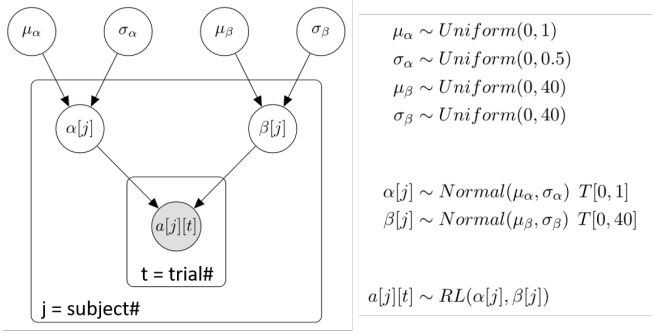

Figure 4: Graphical representation of hierarchical Bayesian Modeling (Gelman et al., 2013). At the group level (top), we sampled the group mean and group standard deviation of α and β parameters from weakly informative priors (uniform and bounded). At the individual level (middle), we sampled α and β for each participant from the group mean and standard deviation just sampled. Individual participants' parameters were used to calculate the likelihood of each action on every trial based on the $\alpha\beta$ model. $T[m,n]$ indicates truncation of distribution. Unfilled circles represented latent variables (in this case, group and individual level model parameters); the filled circle represented observed variable (in this case, participants' choices on each trial).

Fitting procedure We used computational modeling and model comparison to probe trial-by-trial learning dynamics. We fitted five RL models (see Computational models) using hierarchical Bayesian Methods (Gelman et al., 2013) jointly to all participants, instead of to each participant independently. Compared to the classic participant-wise maximum likelihood estimation approach, hierarchical model fitting provides better point estimates for individual participants and allows inference of effects on parameters at the group level (Katahira, 2016). We used a state-of-the-art Markov Chain Monte Carlo (MCMC) sampling, no-U-Turn sampler, implemented in the probabilistic programming language STAN (Carpenter et al., 2017), to sample from the joint posterior distribution. The empirical distribution of the samples approximates the true posterior, which additionally provides a measure of uncertainty, besides point estimates of individual model parameters, allowing more robust statistical inference.

We use the simplest model, $\alpha\beta$, as an example to describe the procedure (Fig. 4). We first sampled group-level parameters, including means and standard deviations for learning rate α (μ_α and σ_α) and inverse temperature β (μ_β and σ_β), from weakly informative priors (uniform and bounded). We then sampled parameters for each participant using these group level parameters: for example, $\alpha[j]$ for participant j was sampled from a normal distribution, $Normal(\mu_\alpha, \sigma_\alpha)$. Note that since the α parameter should be constrained to $[0,1]$, we truncated this normal distribution accordingly. The individual participants' parameters were then used to calculate 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.

For each model, we ran 4 MCMC chains in parallel, with each chain generating 5000 samples (2500 warmup samples), resulting in 10000 samples per model for later inference. We checked the convergence of all models using *bbstanlib* (Baribault, 2019). In particular, none of the models generated any divergent transitions during sampling; $\hat{R}$ for all free parameters were below 1.05; and the effective sample size for all free parameters were more than 200.

Table 1: WAIC scores

Model	$\alpha\beta$	$\alpha^+0\beta$	$\alpha^+\alpha^-\beta$	$\alpha^+0\beta f$	$\alpha^+\alpha^-\beta f$
WAIC	30042	29054	28846	28460	**28337**

Model comparison We performed model comparison at the group level with WAIC (Watanabe, 2013), an information criterion that penalizes model complexity appropriately for hierarchical Bayesian models - smaller WAIC indicates a better fit to the data, controlling for complexity. The $\alpha^+\alpha^-\beta f$ model with asymmetric learning rates and the forgetting parameter had the lowest (best) WAIC score (Table 1).

Our results support recent findings in deterministic learning tasks that including forgetting captured behavior better.

Moreover, the model with α^- fitted as a free parameter had a better WAIC score than simpler models. This confirms that, at the group level, participants did learn from negative feedback.

The group-level mean parameter for α^+ was significantly higher than that for α^- (direct comparison of the empirical distribution of the 10000 posterior samples, $p < 0.0001$). Thus, participants learned much more strongly from positive than negative feedback ($\mu_{\alpha^+} = 0.21$ (95% CI = [0.18, 0.26]); $\mu_{\alpha^-} = 0.004$ (95% CI = [0.0001, 0.01]); Fig. 5AB). The small value of group level mean of α^- reflects the fact that, when each participant was fitted individually, the majority of the participants favored the simpler models without α^-. However, since some participants did learn from negative feedbacks ($\alpha^- > 0$), when all participants were jointly fitted hierarchically, having α^- still improved WAIC.

We validated the best fitting model, $\alpha^+\alpha^-\beta f$, by simulating synthetic choice trajectories from fitted parameters (Fig. 2B) (Palminteri, Wyart, & Koechlin, 2017). Model simulations captured age effects on overall performance (inverse-U shape of performance against age groups; Fig. 2B).

Age differences in model parameters We next investigated which processes drove the changes in performance over age. Specifically, we tested whether parameters of the best fitting model systematically changed with age. To do so, we extended the hierarchical model fitting approach over the previously best fitting model, $\alpha^+\alpha^-\beta f$. Note that directly assessing this relationship by regressing the individual parameters estimated from the previous model against age is not statistically appropriate: the individual parameters were sampled jointly during MCMC, thus violating linear regression assumptions of independent and identical distribution of samples.

The hierarchical Bayesian approach provides a built-in way to test for effects of external variables on parameter models (Gelman et al., 2013). Specifically, we incorporated the regression assumptions into the graphical model of hierarchical Bayesian model, and directly sampled regression coefficients for age jointly with other model parameters. More precisely, for each of the free parameters $\theta, \theta \in \{\alpha^+, \alpha^-, \beta, f\}$, we first sampled an intercept term, $\theta_{intercept}$ for each participant, identical to how we sampled parameters for individual participants from group level parameters (Fig. 4). In addition, we also sampled a linear term, θ_{linear} from a weakly informative prior (uniform and bounded). The parameter $\theta[j]$, used to compute the likelihood of participant j's choice trajectories, was defined as:

$$\theta[j] = \theta_{intercept} + \theta_{linear} * age[j], \quad (4)$$

where $age[j]$ was the z-scored age of participant j, thus implementing a regression directly into the full model.

For quadratic regressions, we additionally sampled a quadratic term, $\theta_{quadratic}$, and $\theta[j]$ became:

$$\theta[j] = \theta_{intercept} + \theta_{linear} * age[j] + \theta_{quadratic} * age[j]^2. \quad (5)$$

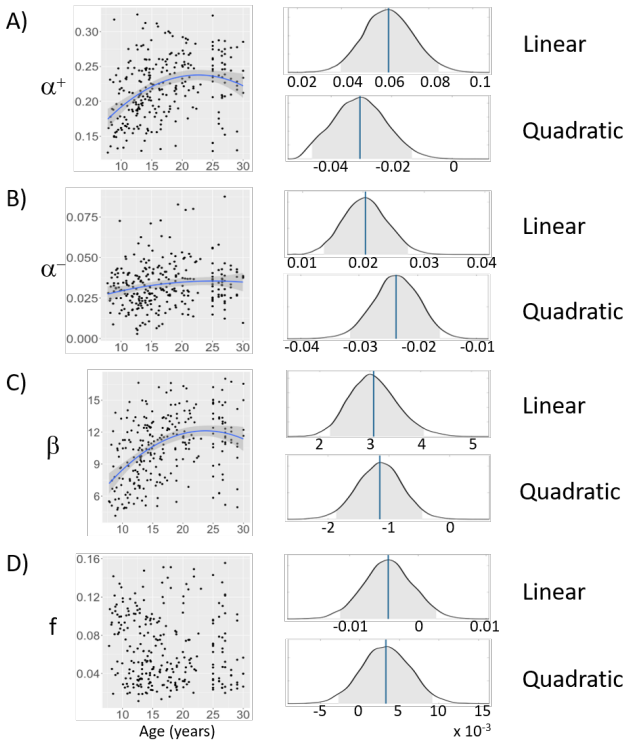

Figure 5: Age effects on model parameters. We directly incorporated age-related parameters into MCMC sampling to test within the hierarchical Bayesian modeling framework whether age had a linear or quadratic effect on all four model parameters: α^+ (A), α^- (B), β (C), f (D). Left panel: individual parameters from the original $\alpha^+\alpha^-\beta f$ model plotted against age. For visualization, we included a quadratic regression line; the shaded region indicates 95% CI (blue curves). Right: distribution of 10000 samples for θ_{linear} (top) and $\theta_{quadratic}$ (bottom). The blue vertical line represents the mean of all samples. Shaded region shows 95% confidence interval. The 10^{-3} scaling applied only to the quadratic effect of age on the forgetting parameter f.

To test for significant linear and/or quadratic effects of age on the model parameters, we examined whether the empirical distribution of all 10000 samples for θ_{linear} and/or $\theta_{quadratic}$ were significantly different from 0 (Fig. 5). We found significant linear and quadratic effects of age on α^+ (linear: $p < 0.0001$; quadratic: $p = 0.0002$), α^- (linear: $p < 0.0001$; quadratic: $p < 0.0001$), and β (linear: $p < 0.0001$; quadratic: $p = 0.001$). We also found that f_{linear} and $f_{quadratic}$ were not significantly different from 0 (linear: $p = 0.88$; quadratic: $p = 0.86$), indicating that there was no effect of age on forgetting rate.

We found linear and quadratic effects of age in parameters α^+, α^- and β (Fig. 5 left). The trajectory of change over age for α^+ and β closely mimicked that for performance, with an inverse U-shape peaking around age 22 and 24 respectively. Moreover, α^+_{linear} was significantly larger than α^-_{linear} ($p < 0.0001$), suggesting that age had a larger effect on sensitivity

to positive over negative feedback, with an increasing bias for positive learning rate(Fig. 5AB left).

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 reacting 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 with adolescence.

At the population level, computational model comparison (Table 1) 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 than negative feedback. This asymmetry has been widely observed in previous studies (Master et al., 2020; van den Bos, Cohen, Kahnt, & Crone, 2012; Hauser, Iannaccone, Walitza, Brandeis, & Brem, 2015), potentially due to differential mechanisms integrating positive and negative feedback (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007). 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 (Master et al., 2020).

Consistent with the age effects observed in previous work using tasks with probabilistic (Eckstein, Master, Dahl, Wilbrecht, & Collins, 2019) and deterministic (Master et al., 2020) feedback, our behavioral and modeling results suggest that learning 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 learning rate parameters (α^+, 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 development are generally consistent with previous work using a deterministic "RLWM' learning task tested in the same participants as data shown here (Master et al., 2020) and a probabilistic task with same the same overall task structure as the Butterfly task used here, but different feedback methods (Davidow et al., 2016).

However, we did not find higher performance in adolescents than adults, as had been observed in this previous Butterfly task study (Davidow et al., 2016) (Fig. 2A, Fig. 3A). Using the same age bin as (Davidow et al., 2016), we found that the performance in 20-30-year-olds was significantly higher than 13-17-year-olds (unpaired t-test, $t(172) = 2.23, p = 0.027$).

The Davidow et al (2016) finding was interpreted as an upside of slower learning that led to more robust integration over time of information, and thus higher overall performance under uncertainty at younger ages. 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. Furthermore, the inverse U-shape itself is dependent on the degree of exploration (Nussenbaum & Hartley, 2019; Davidow et al., 2016; Wilson & Collins, 2019). Learning rates were smaller in our study compared to (Davidow et al., 2016): the group level mean for α^+ in our sample was 0.21, whereas in (Davidow et al., 2016), the mean was around 0.3 and 0.6 for adolescents and adults respectively (Fig. 2B in (Davidow et al., 2016)). Thus, in Davidow et al's higher range of learning rate, 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, but over a different baseline value, leading to opposite effects on performance.

Therefore, while we found a similar trend as in (Davidow et al., 2016) that learning rates increased with age (Fig. 5), 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 feedback phase or differences in socioeconomic status and education level between the samples. For example, it is possible that the peak in performance around age 24 in our sample (Fig. 3A) might be driven by the fact that our 18-25 year-olds are 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 Davidow et al. (2016).

Nevertheless, our results support other previous developmental findings. In particular, we found a decrease in exploration with age (Master et al., 2020; Christakou, Gershman, Niv, & Simmons, 2013), an increase in learning rate previously observed in both deterministic (Master et al., 2020) and probabilistic learning tasks (Eckstein et al., 2019). We also replicated in a probabilistic task a surprising recent finding in a deterministic task (Master et al., 2020): we found no change in forgetting, a component usually attributed to working memory's role in learning, and thus expected to change during adolescence.

In conclusion, we sought to examine the development of learning in a 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 from age 8-24 that were explained by an increase in learning from rewarded outcomes (corresponding to a narrower window of information integration) and a decrease in exploration. Changes in forgetting and learning from non rewarded outcomes varied across subjects but did not show systematic change with development. These data and models help ex-

plain why learning and decision making differ at different stages of development and why a 'one-size-fit-all' approach may not equally serve youth at different stages.

References

Baribault, B. (2019). *bbstanlib: A library of helper functions for stan/matlabstan.*

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. doi: 10.1038/nn1954

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017, January). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1), 1–32. doi: 10.18637/jss.v076.i01

Christakou, A., Gershman, S., Niv, Y., & Simmons, A. (2013). *Neural and Psychological Maturation of Decision-making in Adolescence and Young Adulthood | Journal of Cognitive Neuroscience | MIT Press Journals.*

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024–1035.

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.

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

Eckstein, M., Master, S., Dahl, R., Wilbrecht, L., & Collins, A. (2019). Modeling the development of decision making in volatile environments using strategies, reinforcement learning, and bayesian inference.. doi: 10.32470/CCN.2019.1409-0

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104*(41), 16311–16316. doi: 10.1073/pnas.0706111104

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

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC. doi: 10.1201/b16018

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595.

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. doi: 10.1016/j.neuroimage.2014.09.018

Katahira, K. (2016, August). How hierarchical models improve point estimates of model parameters at the individual level. *Journal of Mathematical Psychology*, *73*. doi: 10.1016/j.jmp.2016.03.007

Larsen, B., & Luna, B. (2018). Adolescence as a neurobiological critical period for the development of higher-order cognition. *Neuroscience & Biobehavioral Reviews*, *94*, 179–195. doi: 10.1016/j.neubiorev.2018.09.005

Master, S. L., Eckstein, M. K., Gotlieb, N., Dahl, R., Wilbrecht, L., & Collins, A. G. E. (2020, February). Distentangling the systems contributing to changes in learning during adolescence. *Developmental Cognitive Neuroscience*, *41*, 100732. doi: 10.1016/j.dcn.2019.100732

Niv, Y. (2009, June). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154. doi: 10.1016/j.jmp.2008.12.005

Nussenbaum, K., & Hartley, C. A. (2019, December). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience*, *40*, 100733. doi: 10.1016/j.dcn.2019.100733

Palminteri, S., Wyart, V., & Koechlin, E. (2017, June). The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences*, *21*(6), 425–433. doi: 10.1016/j.tics.2017.03.011

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.

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. doi: 10.1093/cercor/bhr198

Watanabe, S. (2013). A Widely Applicable Bayesian Information Criterion. , 31.

Wilson, R., & Collins, A. G. E. (2019). *Ten simple rules for the computational modeling of behavioral data | eLife.*