

## **Distangling the systems contributing to changes in learning during adolescence**

Authors: Sarah L. Master, Maria K. Eckstein, Neta Gotlieb, Ronald Dahl, Linda Wilbrecht, Anne G.E. Collins

### **Abstract**

Multiple neurocognitive systems contribute simultaneously to learning. For example, dopamine and basal ganglia (BG) systems are thought to support reinforcement learning (RL) by incrementally updating the value of choices, while the prefrontal cortex (PFC) contributes different computations, such as actively maintaining precise information in working memory (WM). It is commonly thought that WM and PFC show more protracted development than RL and BG systems, yet their contributions are rarely assessed in tandem. Here, we used a simple learning task to test how RL and WM contribute to changes in learning across adolescence. We tested 187 subjects ages 8 to 17 and 53 adults (25-30). Participants learned stimulus-action associations from feedback; the learning load was varied to be within or exceed WM capacity. Participants age 8-12 learned slower than participants age 13-17, and were more sensitive to load. We used computational modeling to estimate subjects' use of WM and RL processes. Surprisingly, we found more robust changes in RL than WM during development. RL learning rate increased significantly with age across adolescence and WM parameters showed more subtle changes, many of them early in adolescence. These results underscore the importance of changes in RL processes for the developmental science of learning.

**Key words (max 6):** development, reinforcement learning, working memory, computational modeling, adolescence

**Highlights:**

- Subjects combine reinforcement learning (RL) and working memory (WM) to learn
- Computational modeling shows RL learning rates grew with age during adolescence
- When load was beyond WM capacity, weaker RL compensated less in younger adolescents
- WM parameters showed subtler and more puberty-related changes
- WM reliance, maintenance, and capacity had separable developmental trajectories
- Underscores importance of RL processes in developmental changes in learning

**Acknowledgements:** We gratefully acknowledge the numerous people who contributed to this research: Lance Kriegsfeld, Celia Ford, Jennifer Pfeifer, Megan M. Johnson, Liyu Xia, Rachel Arsenault, Josephine Christon, Shoshana Edelman, Lucy Eletel, Haley Keglovits, Julie Liu, Justin Morillo, Nithya Rajakumar, Nick Spence, Tanya Smith, Benjamin Tang, Talia Welte, Lucy B. Whitmore, and Amy Zou. We are grateful to our participants and their families for their time and participation in this study.

**Funding:** This work was funded by National Science Foundation SL-CN grant #1640885 to RD, AGEK, and LW.

The authors declare no conflict of interest.

## Introduction

There is increasing evidence that multiple neural systems contribute to human learning ((Hazy, Frank, & O'Reilly, 2007; Myers et al., 2002)), even in simple cognitive paradigms previously modeled with a single learning process ((Bornstein & Daw, 2012; Bornstein & Norman, 2017; Collins & Frank, 2012)). Basal ganglia (BG) dependent reinforcement learning (RL) processes are thought to be supplemented by multiple other systems, including prefrontal executive functions ((Badre, Kayser, & D'Esposito, 2010)) such as working memory (WM; (Collins & Frank, 2012)) and model-based planning ((Daw et al., 2011)), as well as hippocampus-based episodic memory ((Bornstein & Daw, 2012; Bornstein & Norman, 2017; Davidow et al., 2016; Myers et al., 2002; Wimmer et al., 2014)). To understand developmental changes in learning, it is important to carefully capture the contributions of these multiple systems to learning. Previous work has shown differential developmental trajectories for RL, episodic memory, and model-based planning (Crone, et al., 2006; Decker et al., 2016; Selmeczy et al., 2018; Somerville, Jones, & Casey, 2010). Here, we investigate how the relative contributions of RL and WM change during development.

RL is an incremental learning process which updates stored choice values from the discrepancy between obtained and expected reward (the reward prediction error; RPE) in order to maximize future rewards. Sensitivity to rewarding outcomes in the ventral striatum has been shown to peak during mid-adolescence, relative to children and adults ((Galvan et al., 2006; Braams et al., 2015; Gulley et al., 2018; Somerville et al., 2010)). It has been proposed that the BG are 'mature' by mid adolescence, but it is also known that structures that provide inputs to the BG continue to show anatomical and functional change ((Braams, van Duijvenvoorde, Peper, & Crone, 2015; Casey et al., 2008; A. Galvan et al., 2006a; Galvan, Delevich, & Wilbrecht, 2019; Van Den Bos, Cohen, Kahnt, & Crone, 2012)).

The prefrontal cortex (PFC) is known to develop late into adolescence and the third decade of life ((Casey et al., 2008; Giedd, Blumenthal, & Jeffries, 1999; Larsen & Luna, 2018)) and is thought to be critical for WM performance ((Curtis & D'Esposito, 2003; Miller & Cohen, 2001)). WM allows for the immediate and accurate storage of information, but representations in WM are thought to decay quickly with time and are

subject to interference, as the information that can be held in WM is limited ((D'Esposito & Postle, 2015; Oberauer et al., 2018)). Therefore there may be tradeoffs to using fast capacity-limited WM, versus slower, capacity unlimited RL-based learning systems.

Behavioral testing and computational modeling can be used to disentangle the roles of RL and WM in human learning, and how their use differs between individuals. Using a deterministic reward-learning task called “RLWM” that taxes WM by varying the amount of information to learn in each block, we have previously isolated contributions of WM and RL learning ((Collins, 2018; Collins & Frank, 2012, 2018)). In multiple studies we found that participants mainly used WM for learning when the load was within their capacity, and otherwise compensated with RL. We also found that learning deficits in schizophrenia were a result of weakened WM contributions with intact RL ((Collins & Frank, 2014; Collins et al., 2017). Here we used the same approach to investigate the maturation of WM and RL and their relative contribution to learning across adolescent development (sampling subjects 8-17 and 25-30).

Using behavioral testing and computational modeling, we examined three separate hypotheses of how RL- and WM-based learning develop relative to each other. Our first hypothesis was that both RL and WM systems' contributions to learning show continued development into later adolescence such that both systems are dynamic throughout the teenage years. Developmental changes in WM are prominent in the literature on development of executive function, including the maintenance ((Geier et al., 2009; McAuley & White, 2011)) and manipulation of information in WM ((Crone et al., 2006; Huizinga, Dolan, & van der Molen, 2006)), as well as the precision of the representations in WM ((Luna, 2009)). There is also a strong literature showing changes in RL learning systems from childhood to adulthood, such as dynamic changes in reward sensitivity in the striatum across adolescence ((Braams et al., 2015; Cohen et al., 2010; Davidow et al., 2016; Somerville et al., 2010)) and changes in learning rate between adolescence and adulthood ((Davidow et al., 2016)).

Our second hypothesis emphasized the relative importance of WM development over that of RL in accounting for changes in learning in adolescence. Dual systems models and other popular models of adolescent development place great weight on the late maturation of the PFC and PFC-dependent executive functions, such as WM or

model-based learning ((Casey et al., 2008; Decker et al., 2016; Huizinga et al., 2006; Steinberg, 2005)). Our second hypothesis, therefore predicts that even though RL may be developing, we should observe stronger effects of age on WM, which is PFC-based, and WM changes should be the primary driver of changes in learning through adolescence.

Finally, we hypothesized that pubertal onset may significantly impact WM processes. There is growing evidence that gonadal hormones affect inhibitory neurotransmission and other variables in the prefrontal cortex of rodents (Delevich, Piekarski, & Wilbrecht, 2019; Delevich, Thomas, & Wilbrecht, 2019; Juraska & Willing, 2017; Piekarski, Boivin, & Wilbrecht, 2017; Piekarski, Johnson, et al., 2017). We therefore predicted that WM parameters would differ in children with different pubertal status or gonadal hormone concentration.

To evaluate these hypotheses, we tested children and adolescents aged 8 to 17 years old and adults aged 25 to 30 years old on the RLWM task (Collins & Frank, 2012). We then fit computational models of behavior to subjects' performance and assessed how these parameters changed with age, pubertal development and salivary testosterone levels. Using these established methods to disentangle the contributions of RL and WM, we found changes in RL spanning adolescent development, but much weaker changes in WM contributions. WM differences did show relationships with pubertal variables.

Overall, these data support the somewhat surprising conclusion that changes in RL systems are important drivers of change in simple associative learning during adolescence. The results also support further inquiry into the role of pubertal processes in WM function.

## **Methods**

**Subject testing.** All procedures were approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley. After entering the testing room, subjects under 18 years old and their guardians provided their informed assent or permission. All guardians were asked to fill out a demographic form, and guardians of children aged 10 or younger completed an additional pubertal development questionnaire about their child. Subjects were led into a quiet testing room in view of their guardians,

where they used a video game controller to complete four computerized tasks. An hour after the start of the experimental session and in between tasks, subjects provided a 1.8 mL saliva sample. At the conclusion of the tasks, subjects were asked to complete a short questionnaire which assessed their pubertal development ((Petersen & Crockett, 1988)) and collected other basic information, like their height and weight. Subjects under 10 years old did not complete the pubertal development questionnaire. Subjects were then compensated with one \$25 Amazon gift card for their time.

Participants over 18 provided informed consent and completed all forms themselves. They also answered retroactive questions about puberty, otherwise all testing procedures were identical.

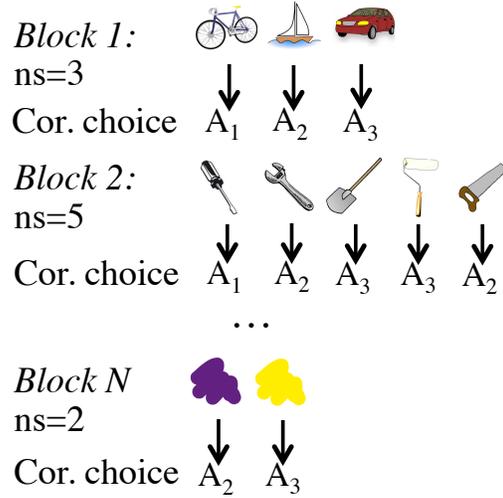
**Demographics.** We recruited 191 children and 55 adults to participate in this study. Out of those who reported their race, 60 subjects identified as Asian, 10 African-American, and 6 Native American or Pacific Islander. 28 were of mixed race. The remaining 127 subjects identified as Caucasian. 29 subjects identified as Hispanic. 4 children and 1 adult failed to complete the task, either out of disinterest or because of controller malfunction. All subjects who completed the task performed above chance (33% accuracy), so none were excluded for poor task performance. 187 children (89 female, mean (std) age 12.62 (2.76) years) and 54 adults (28 female, mean (std) age 26.77 (1.49) years) were included in analyses.

**Experimental paradigm.** The experiment described in this work was the second of four administered tasks. The preceding task was a 5 to 10 minute deterministic learning task in which subjects learned to select one correct action out of four possible actions. Halfway through, the correct action switched. The results for this and the other tasks will be reported elsewhere.

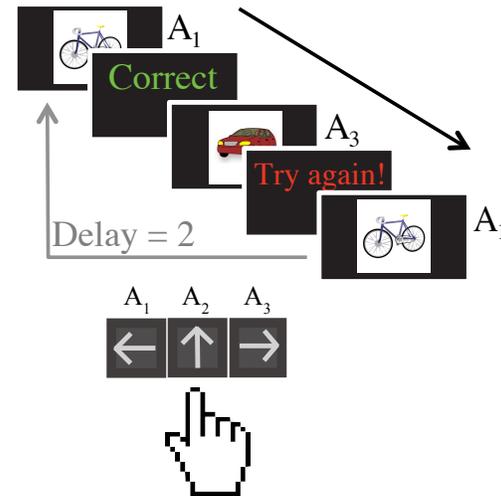
This experiment is based on the “RLWM” task described in ((Collins, Ciullo, Frank, & Badre, 2017; Collins & Frank, 2012, 2018)), adapted for the developmental population. To make the task more engaging for children, participants were told that aliens had landed on earth and wanted to teach us their alien language. In this alien language, one button on the controller was matched with each image on the screen.

Participants completed one block of training and then ten independent learning blocks, for a total duration of less than 25 minutes (mean duration 16.9 minutes, range 14 – 25 minutes).

### Protocol



### Block 1: 3 trial example



**Figure 1: Experimental protocol.** In each block, participants learned to select the correct action for each image for a new set of images. At the beginning of each block, the full set of images was presented for familiarization, then single trials began. On each trial of a block, participants responded to each stimulus by pressing one of three buttons on a hand-held controller. Immediately after responding they received deterministic truthful feedback (Correct/Try again!), before moving on to the next trial after a fixed interval. Crucially, participants needed to learn different numbers of stimuli (set size) in different blocks. Set size ( $ns$ ) varied from 2 to 5.

In each block, subjects were presented with a new set of visual stimuli of set size  $ns$ , with set sizes between  $ns=2$  and  $ns=5$ . Each visual stimulus was presented 12-14 times in a pseudo-randomly interleaved manner (controlling for a uniform distribution of delay between two successive presentations of the same stimulus within  $[1:2*ns]$  trials, and the number of presentations of each stimulus), for a total of  $ns*13$  trials. At each trial, a stimulus was presented centrally on a black background (Figure 1). Subjects had up to 7 seconds to answer by pressing one of three buttons on the controller. Key press was followed by visual feedback presentation for 0.75 seconds, then a fixation period of 0.5 seconds before the onset of the next trial. For each image, the correct key press was the same for the entire block, and all feedback was truthful. Upon pressing the correct key

subjects were shown “Correct” while pressing any other key led to “Try again!” Failure to answer within 7 seconds was indicated by a “No valid answer” message. Stimuli in a given block were all from a single category of familiar, easily identifiable images (e.g. colors, fruits, animals) and did not repeat across blocks. Participants were shown all stimuli at the beginning of each block, and encouraged to familiarize themselves with them prior to starting the learning block.

We varied the set size across blocks: out of 10 blocks, 3 had set size  $ns=2$ , 3 set size  $ns=3$ , 2 set size  $ns=4$ , and 2 set size  $ns=5$ . We have shown in previous work that varying set size provides a way to investigate the contributions of capacity- and resource-limited working memory to reinforcement learning.

**Saliva collection.** In addition to self-report measures of pubertal development, we also collected saliva from each of our subjects to quantify salivary testosterone. Testosterone is a reliable measure of pubertal status in boys and girls and is associated with changes in brain and cognition in adolescence ((Herting et al., 2014; Peper et al., 2011)).

Subjects refrained from eating, drinking, or chewing anything at least an hour before saliva collection. Subjects were asked to rinse their mouth out with water approximately 15 minutes into the session. At least one hour into the testing session, they were asked to provide 1.8 mL of saliva through a plastic straw into a 2 mL tube. Subjects were instructed to limit air bubbles in the sample by passively drooling into the tube, not spitting. Subjects were allotted 15 minutes to provide the sample. After the subjects provided 1.8 mL of saliva, or 15 minutes had passed, the sample was immediately stored in a  $-20^{\circ}\text{F}$  freezer. The date and time was noted by the experimenter. The subjects then filled out a questionnaire of information which might affect the hormones measured in the sample (i.e. whether the subjects had recently exercised).

All subjects were asked whether they would like to complete two more saliva samples at home for additional compensation (another \$25 Amazon gift card). Subjects who agreed to do the optional saliva samples were sent home with two 2 mL tubes, straws, and questionnaires identical to the one completed in lab. They were asked to complete each sample between 7:00 and 9:00 am on two different days following the testing session. Subjects were asked to refrain from eating, drinking, or brushing their

teeth before doing the sample, and to fill out each questionnaire as soon as they were finished collecting saliva, taking care to note the date and time of the sample. Subjects were also instructed to keep the samples in the freezer, then wrap them in a provided ice pack in order to deliver the samples to the lab. Once both samples were complete, subjects contacted an experimenter and scheduled a time to return the samples, who gave the subjects their additional compensation, took note of any abnormalities in the samples, and immediately stored them in a -20 degree freezer. Samples in the -20 degree freezer were transferred weekly to a -80 degree freezer in an adjacent facility.

**Salivary Testosterone Testing.** Salivary testosterone was quantified using the Salimetrics Salivary Testosterone ELISA (cat. no. 1-2402, Bethesda, MA). Intra- and inter-assay variability for testosterone were 3.9% and 3.4%, respectively. Samples below the detectable range of the assay were assigned a value of 5 pg/mL, 1 pg below the lowest detectable value. Final testosterone sample concentration data were cleaned with a method developed by Shirtcliff and Byrne (in prep). Specifically, we produced a mean testosterone concentration from every salivary sample obtained from each of our subjects (every subject provided 1 to 3 samples). Subjects who had multiple samples below the detectable range of the assay (6 pg/mL) had their mean testosterone concentration replaced with 1 pg below the lowest detectable value (5). There were no subjects with any samples above the detectable range. Within subjects aged 8 to 17 only, outliers greater than 3 standard deviations above the group mean were fixed to that value, then incremented in values of +0.01 to retain the ordinality of the outliers.

**Model-independent analyses.** We now describe how we analyzed the data from the behavioral task. To assess learning, we calculated the proportion of correct trials for each subject on each stimulus iteration. Each individual stimulus was repeated 13 to 14 times within a block. Within each set size, we calculated each subject's average percentage of correct responses at each stimulus iteration (Figure 2a).

We quantified the effect of set size on performance by calculating a set size slope (Figure 2b). The set size slope was a linear contrast of the form:

$$-1.5 * perf(ns5) - 0.5 * perf(ns4) + 0.5 * perf(ns3) + 1.5 * perf(ns2)$$

where  $perf(ns)$  is the average overall performance of trials within a block of set size  $ns$ .

Set size is correlated with delay between repetitions of the same stimulus, thus the previous measure may capture two separate aspects of the protocol that affect working memory (load and decay). In order to more precisely assess the effects of our set size manipulation on performance, we used a logistic regression to model trial-by-trial accuracy as a function of previous correct trials ( $pcor$ ), previous incorrect trials ( $pinc$ ), number of trials since the last presentation of the same stimulus ( $delay$ ), and set size ( $ns$ ) as predictors (Collins & Frank, 2012).

To understand the effects of development on behavior and on model parameters, we ran a series of analyses at group and individual levels. First, following the practice in the literature to separate “children” from “adolescents” and adults (Potter, Bryce, & Hartley, 2017) we grouped 8 to 12, 13 to 17, and 25 to 30-year-olds into separate groups and ran one-way ANOVAs on behavioral and modeling measures of interest, looking for broad group-based, age-related differences in cognition; ANOVAs were replaced by Kruskal-Wallis non-parametric tests for non-normal measurements. Where group effects were present, we ran post-hoc tests to identify which group drove the effect. We used  $t$ -tests when the measure was normally distributed, rank tests otherwise.

While helpful to visualize results, the binning into coarse 8 to 12 and 13 to 17-year-old groups for non-adult participants was arbitrary. We also analyzed the data of this 8 to 17-year-old subsample as a function of age as a continuous measure. Specifically, we ran non-parametric (Spearman) correlation tests on behavioral measures with age as a continuous predictor within the non-adult group.

Additionally, to further investigate the effects of age and pubertal development on our outcome measures, we examined relationships between behavior and model parameters as a function of pubertal development score (PDS) and salivary testosterone, either as continuous predictors, or defining binned groups (see supplementary material). Given that girls tend to begin puberty earlier than their male peers, and that the production of testosterone tracks the development of boys and girls differently, we also

ran analyses of pubertal effects separating male and female subjects, and combining them. We also grouped participants aged 8 to 17 into narrower age bins based on quartiles. Further descriptions of how we grouped subjects into age, PDS, and testosterone bins, plus additional statistical test methods and results, are included in the supplemental material.

### **Computational modeling**

We used computational modeling to fit subject behavior and better quantify the separate involvements of reinforcement learning and working memory in task execution. We have shown that our model is able to separate the contributions of these two learning systems in both general ((Collins, 2018; Collins & Frank, 2012, 2018)) and specific populations ((Gold et al., 2017)). We tested 6 candidate models all built upon a simple reinforcement learning (RL) algorithm.

**Classic RL.** The simplest model is a two parameter Q-learner (*RL*), which updates the learned value  $Q$  for the selected action  $a$  given the stimulus  $s$  upon observing each trial's reward outcome  $r_t$  (1 for correct, 0 for incorrect):

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \delta_t$$

where  $\delta_t = r_t - Q_t(s, a)$  is the prediction error, and  $\alpha$  is the learning rate, which is a free parameter. This model is similar to standard RL models as described in Sutton and Barto's Reinforcement Learning: An Introduction ((2017)). Choices are generated probabilistically with greater likelihood of selecting actions that have higher  $Q$ -values. This choice is driven by a softmax choice policy, which defines the rule for choosing actions in response to a stimulus:

$$p(a|s) = \exp(\beta Q(s, a)) / \sum_i (\exp(\beta Q(s, a)))$$

Here,  $\beta$  is the inverse temperature parameter determining the degree to which differences in Q-values are translated into more deterministic choice, and the sum is over the three possible actions  $a_i$ .

**RL with undirected noise (RL $\epsilon$ ).** While the softmax allows for some stochasticity in choice, we also tested a model which allowed for “slips” of action. This was captured in an undirected noise parameter,  $\epsilon$ . Given a model’s policy  $\pi = p(a|s)$ , adding undirected noise consists in defining the new mixture choice policy:

$$\pi' = (1 - \epsilon)\pi + \epsilon U,$$

where  $U$  is the uniform random policy ( $U(a) = 1/n_A$ , with number of actions  $n_A = 3$ ).  $\epsilon$  is a free parameter constrained to values between 0 and 1. This undirected noise captures a choice policy where with probability  $1 - \epsilon$  the agent chooses an action based on the softmax probability, and with probability  $\epsilon$  lapses and chooses randomly. Failing to account for this irreducible noise can allow model fits to be unduly influenced by rare odd data points, like those that may arise from attentional lapses (Nassar & Frank, 2016).

**RL with perseverance (RL $p$ ).** To allow for potential neglect of negative feedback, we estimate a perseverance parameter *pers* such that for negative prediction errors ( $\delta < 0$ ), the learning rate  $\alpha$  is reduced by  $\alpha = (1 - pers)\alpha$ . Thus, values of *pers* near 1 indicate perseverance with complete neglect of negative feedback, whereas values near 0 indicate equal learning from negative and positive feedback.

**RL with forgetting (RL $f$ ).** In this model we allow for potential forgetting of Q-values on each trial, implemented as a decay at each trial toward the initial, uninformed Q-value  $Q_0$ :

$$Q_{t+1} = Q_t + \varphi(Q_0 - Q_t)$$

where  $0 < \varphi < 1$  is the forgetting parameter and  $Q_0 = 1/n_A$ .

**RL with 4 learning rates (RL4).** To improve the fit within the “RL only” class of models, we tested a version of the Q-learner that included a different learning rate  $\alpha$  for each set size. Theoretically, this model could capture set size effects if they were driven by slower learning in higher set sizes.

**RL and working memory (RLWM).** This model incorporates two separate mechanisms by which learning can take place which interact at the level of choice. The first mechanism is a RL model as described above, with an inverse temperature parameter  $\beta$ , learning rate  $\alpha$ , and undirected noise  $\varepsilon$ . The second mechanism is a working memory module.

The WM module stores weights between stimuli and actions,  $W(s,a)$ , which are initialized similarly to RL  $Q$ -values. To model fast storage of information, we assume perfect retention of the previous trial’s information, such that  $W_{t+1}(s_t, a_t) = r_t$ , but include capacity limitation such that at most  $K$  stimuli can be remembered. To model delay-sensitive aspects of working memory (where active maintenance is increasingly likely to fail with intervening time and other stimuli), we assume that WM weights decay at each trial according to  $W_{t+1} = W_t + \varphi_{WM}(W_0 - W_t)$ . The WM policy uses these weights in a softmax choice with added undirected noise, using the same noise parameters as the RL module.

Note that this implementation assumes that information stored for each stimulus in working memory pertains to action-outcome associations. Furthermore, this implementation is an approximation of working memory that assumes 1) rapid and accurate encoding of information when low amounts of information are to be stored; 2) decrease in the likelihood of storing or maintaining items when more information is presented or when distractors are presented; 3) decay due to forgetting. We believe these are key aspects of working memory.

RL and WM involvement in choice is modeled with a WM weight parameterized by the working memory capacity parameter  $K$ , and a WM confidence prior  $\rho$ . The overall choice policy is defined as a mixture using WM weight  $W_{WM} = \rho(\min(1, K/ns))$ :

$$P(a|s) = W_{WM}P_{WM}(a|s) + (1 - W_{WM})P_{RL}(a|s).$$

$\rho$  captures the subject's overall propensity to use WM vs. RL when within WM's capacity. The WM weight then considers the capacity limit of the WM module as indicated by the proportion of items that can be maintained in working memory ( $\min(1, K/ns)$ ) as well as the subjects' prior for relying on WM.

The strongest model in model comparison (by AIC score, figure 3b) was the RLWM model with six free parameters: the RL learning rate  $\alpha$ , WM capacity  $K$ , WM decay  $\varphi$ , WM prior weight  $\rho$ , perseveration parameter *pers*, and undirected noise  $\varepsilon$ . The inverse temperature parameter  $\beta$  was fixed to 100, per previous experience showing that it allows best characterization of WM and RL parameters (Collins, 2018).

**Model fitting procedure.** We used the Matlab constrained optimization function `fmincon` to fit parameters (the Mathworks Inc., Natick, Massachusetts, USA). This was iterated with 20 randomly chosen starting points, to increase the likelihood of finding a global rather than local optimum. All parameters were fit with constraints [0 1], except the capacity parameter  $K$ . Due to the non-continuous nature of  $K$ , each set of random starting points was paired with each of the possible fixed values [2 3 4 5] of  $K$ . The best fit within those possible values of  $K$  was selected as a proxy for optimizing  $K$  alongside the other parameters.

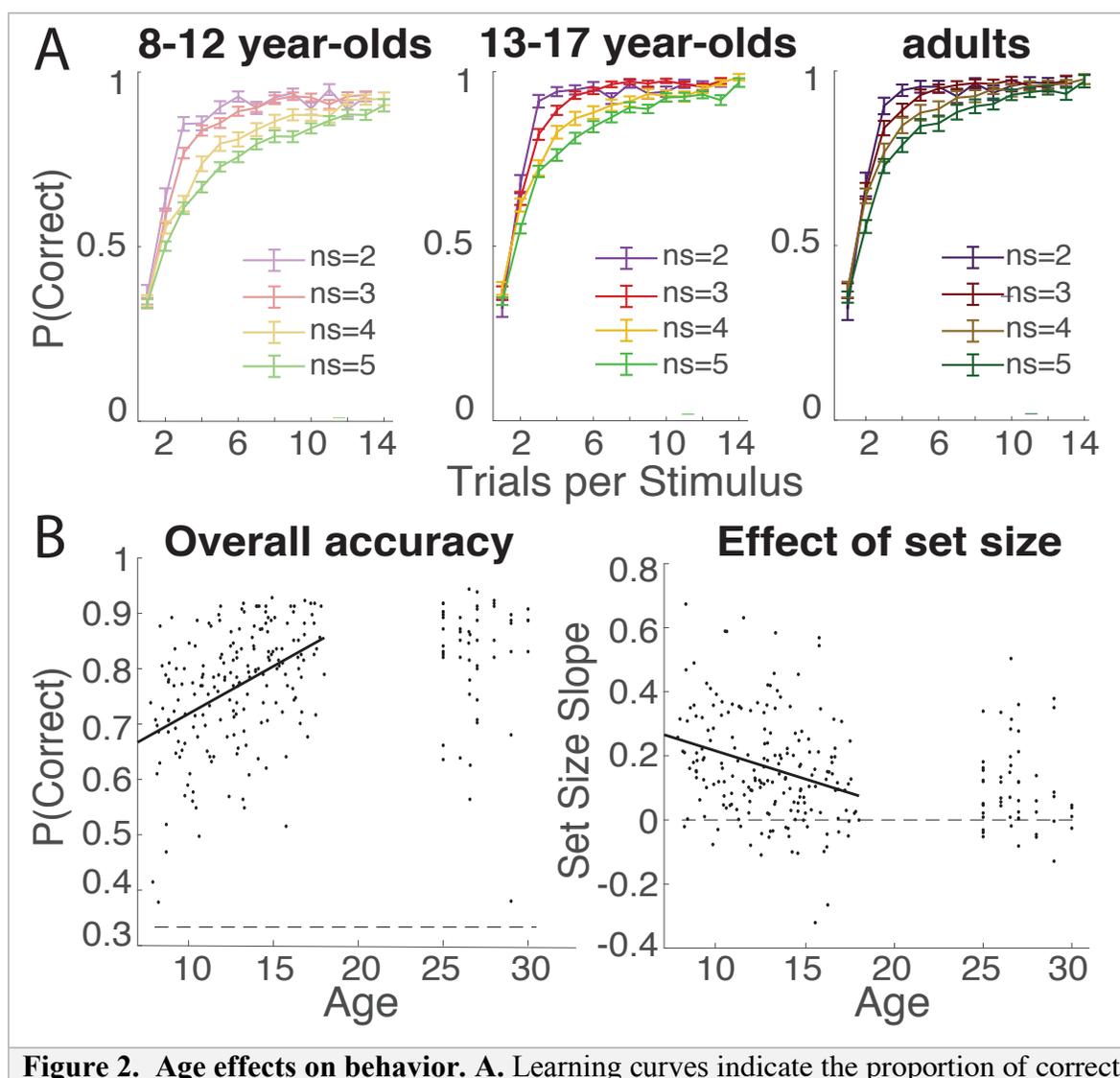
**Model comparison.** We used the Akaike Information Criterion (AIC; Burnham & Anderson, 2002) to assess relative model fits and penalize model complexity. We previously showed that in the case of the RLWM model and its variants, AIC is a better approximation of model fit than Bayesian Information Criterion (BIC; Schwarz, 1978) at recovering the true model from generative simulations (Collins & Frank, 2012). Comparing RLWM and each of the variants of the simple RL model showed that RLWM provided a better fit to the data despite its additional complexity.

**Model simulation.** Model comparison alone is insufficient to assess whether the best fitting model sufficiently captures the data, as it provides only a relative measure of model fit ((Wilson & Collins, submitted)). To test whether our models capture the key

aspects of the behavior (i.e. learning curves), we simulated each model with fit parameters from each subject, with 100 repetitions per subject averaged to represent each subject's contribution to group-level behavioral effects (Figure 3C).

## Results

**Overall accuracy.** To analyze coarse age effects on behavioral and modeling measures, we first grouped participants into three groups by age (8 to 12-year-olds, 13 to 17-year-olds, and 25 to 30-year-olds), and tested the continuous effect of age on performance for non-adult participants.



**Figure 2. Age effects on behavior.** A. Learning curves indicate the proportion of correct

trials as a function of the number of encounters with given stimuli by set size (ns) in subjects aged 8 to 12, 13 to 17, and 25 to 30. All subjects quickly reached asymptotic accuracy in set sizes 2 and 3. In set sizes 4 and 5, learning was more graded. 8-12-year-olds appeared to learn slower than both 13-17-year-olds and adults (25-30). 13-17-year-olds and 25-30-year-olds exhibited similar learning. **B.** Overall accuracy and set size slope (linear effect of set size on performance; see methods) as a function of age. Each dot represents a single subject; lines indicate the best fit in a regression model. In subjects aged 8 to 17 overall accuracy significantly increased with age while set size effect significantly decreased, though it remained positive in all age groups.

All participants performed significantly better than chance (33%, Figure 2b). The mean accuracy for all subjects was 77.91% (median 80%). An ANOVA by age group revealed a main effect of group ( $F(240) = 20.29$ ,  $p = 9.10e^{-9}$ ). Post-hoc t-tests show that this main effect was driven by the differences in performance of the 8-12 year-old group from the 13-17 year-old group ( $t(187) = 5.2$ ,  $p < 10e^{-4}$ ) and the 25-30 year-old group ( $t(152) = 5.2$ ,  $p < 10e^{-4}$ ). 8 to 12-year-olds' performance was significantly worse overall (73% accuracy), while 13 to 17-year-olds and 25 to 30-year-olds performed similarly well (at 80.1% and 82.4% accuracy, respectively;  $t(141) = 1$ ,  $p = 0.3$ ). Within subjects aged 8 to 17, there was a positive correlation of age and overall accuracy (Figure 2b; Spearman  $\rho = 0.44$ ,  $p = 2.e-10$ ), confirming an expected improvement in learning performance with age.

**Learning.** All age groups showed a similar qualitative pattern of learning whereby learning was faster in lower set sizes, a characteristic of combined working memory and reinforcement contributions to learning (Figure 2a). Learning reached asymptote for set sizes 2 and 3 within the first three or four trials, and reached asymptote incrementally in set sizes 4 and 5. There was a strong negative effect of set size on performance (set size slope; see methods;  $t(242) = 14.7$ ,  $p = 2e-35$ ; 207 of 243 participants). An ANOVA on the effect of set size revealed a main effect of age group ( $F(240) = 7.6$ ,  $p = 0.0006$ ). Post-hoc t-tests confirmed that this was driven by a larger negative effect of set size in 8 to 12-year-olds' performance than in 13 to 17-year-olds' ( $t(187) = 2.85$ ,  $p = 0.0049$ ) and adults' ( $t(152) = 3.59$ ,  $p = 0.0004$ ). 13 to 17-year-olds were not more or less affected by set size than 25 to 30-year-olds ( $t(141) = 0.98$ ,  $p = 0.33$ ). Within subjects aged 8 to 17, there was a negative correlation between age and set size slope ( $\rho = -0.28$ ,  $p = 8e-5$ ; Figure 2b),

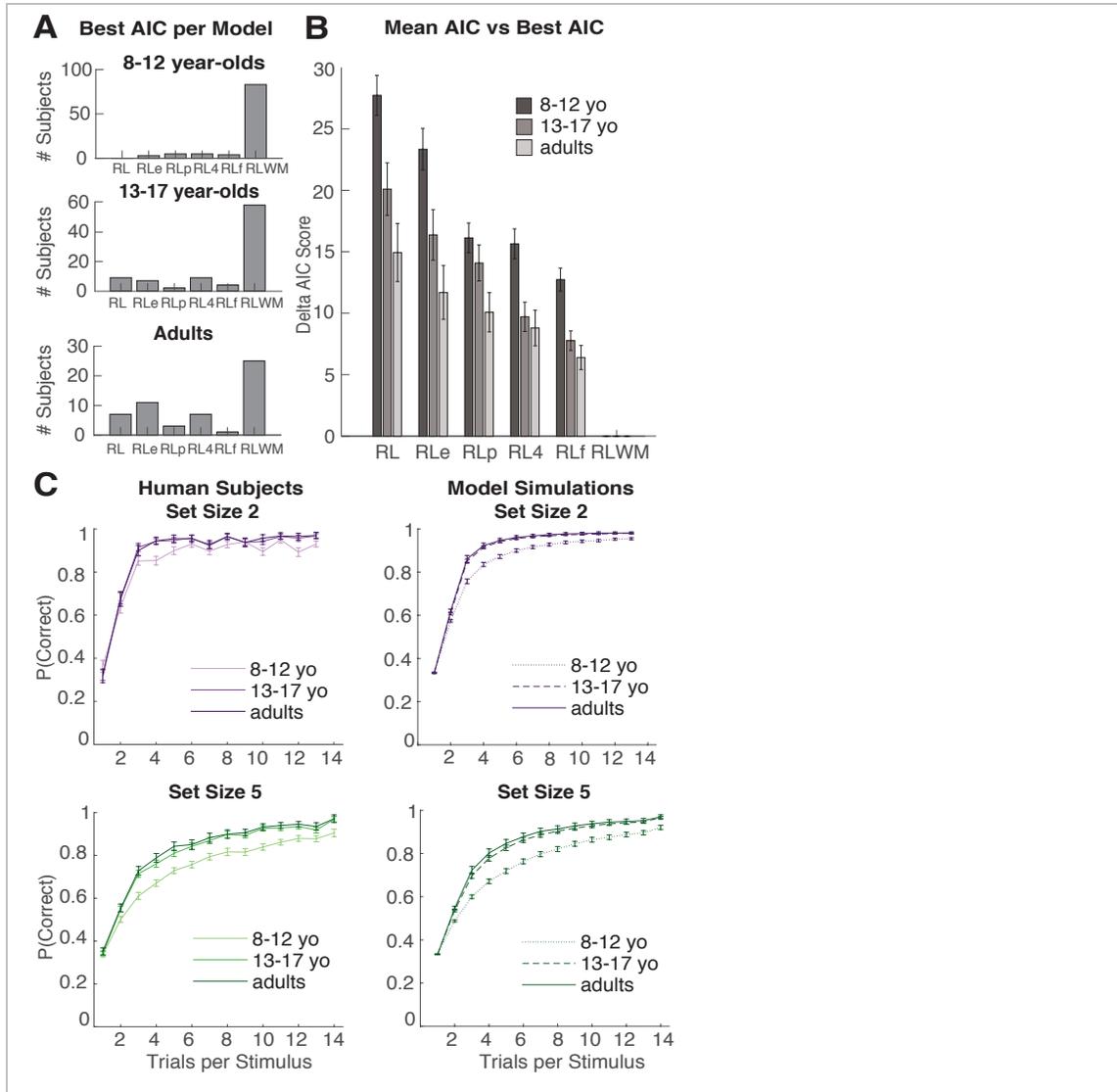
supporting the previous analysis that participants' learning became less sensitive to set size with increased age.

**Reaction time.** Overall, mean reaction time (RT) decreased as a function of age ( $\rho = -0.34$ ,  $p = 2e-6$ ). RT variance also decreased as a function of age ( $\rho = -0.27$ ,  $p = 0.0002$ ). Reaction times were slower on high set size trials ( $t(242) = 33.1$ ,  $p = 0$ ; 242 out of 243 participants). There was no main effect of age group on the set size RT effect ( $F(240) = 1.24$ ,  $p = 0.29$ ).

All age groups and subjects appeared to be sensitive to the set size manipulation in both accuracy and reaction time, supporting the fact that all ages used both RL and WM to learn in this protocol. 8- to 12-year-olds were slower, less accurate, and more sensitive to the set-size manipulation than participants aged 13 and up. Performance rapidly reached adult levels by age 13 and remained stable until age 30. We next sought to use statistical and computational models to better characterize the underlying processes that drove the developmental changes in behavior.

**Logistic regression.** For each subject, we ran a trial-by-trial logistic regression predicting response accuracy with predictors set size, delay since the current stimulus was last presented (two potential markers of WM), previous correct trials for that stimulus, and previous incorrect trials (two potential markers of RL; see methods). Set size had a negative effect on performance in all age groups (8 to 12-year-olds  $t = -6.96$ ,  $p < 0.0003$ ; 13 to 17-year-olds  $t = -4.39$ ,  $p < 0.0003$ ; 25 to 30-year-olds  $t = -3.88$ ,  $p < 0.0003$ ). Delay also had a negative effect on performance in all groups (8 to 12-year-olds  $t = -4.56$ ,  $p < 0.0001$ ; 13 to 17-year-olds  $t = -4.39$ ,  $p < 0.0001$ ; 25 to 30-year-olds  $t = -2.3$ ,  $p = 0.026$ ). Number of previous correct trials had a positive effect on performance in all age groups (8 to 12-year-olds  $t = 21.3$ ,  $p < 0.0001$ ; 13 to 17-year-olds  $t = 8.5$ ,  $p < 0.0001$ ; 25 to 30-year-olds  $t = 8.5$ ,  $p < 0.0001$ ). Number of previous incorrect trials had a negative effect in both children and teenagers but not in adults (8 to 12-year-olds  $t = -11.1$ ,  $p < 0.0001$ ; 13 to 17-year-olds  $t = -4.93$ ,  $p < 0.0001$ ; 25 to 30-year-olds  $t = -1.1$ ,  $p = 0.27$ ).

We tested the relationship between age and the individual logistic regression weights in a multiple regression predicting age from each individual's five regression weights. After excluding subjects with weights 2 standard deviations above or below the mean (10 8 to 17-year-olds), we found that the weight of the fixed effect ( $t(173) = 7.1, p < 10e-4$ ), and the weight of previous correct trials ( $t(173) = -2.9, p = 0.004$ ) predicted age. None of the other predictors (pinc, delay, ns) predicted age ( $p$ 's  $> 0.09$ ).



**Figure 3. Model validation.** **A.** Best fit model per subject. All subjects' behavior was fit with 6 candidate models: reinforcement learning (RL), RL with an epsilon noise parameter (RLe), RL with perseveration (RLp), RL with four learning rates (RL4), RL with forgetting (RLf), and RL with working memory (RLWM; see methods). Plotted here is the number of subjects best fit by each candidate model in each age group as measured with AIC score. RLWM was the best-fitting model for a majority of subjects within subjects aged 8-12, 13-17, and 25-30. **B.** Difference in mean AIC score from the

best fitting model (RLWM). Within subjects 8 to 12 years old (yo), 13 to 17 years old, and adults (25 to 30 yo), we calculated the mean AIC score for each candidate model, then compared to the mean AIC for the winning model (RLWM). Lower numbers indicate better fits. Error bars made with standard error of the mean. **C. Model validation.** Learning curves for participants (left) and model simulations (right) for set sizes 2 and 5 (see SI for all set sizes). RLWM model simulations with individual fit parameters accounted for behavior.

Results so far confirmed our prediction that younger participants would perform worse than older participants. There was evidence for successful RL recruitment in all subjects based on use of previous correct feedback, as well as evidence of WM recruitment based on set-size and delay effects. However, results so far were ambiguous as to whether

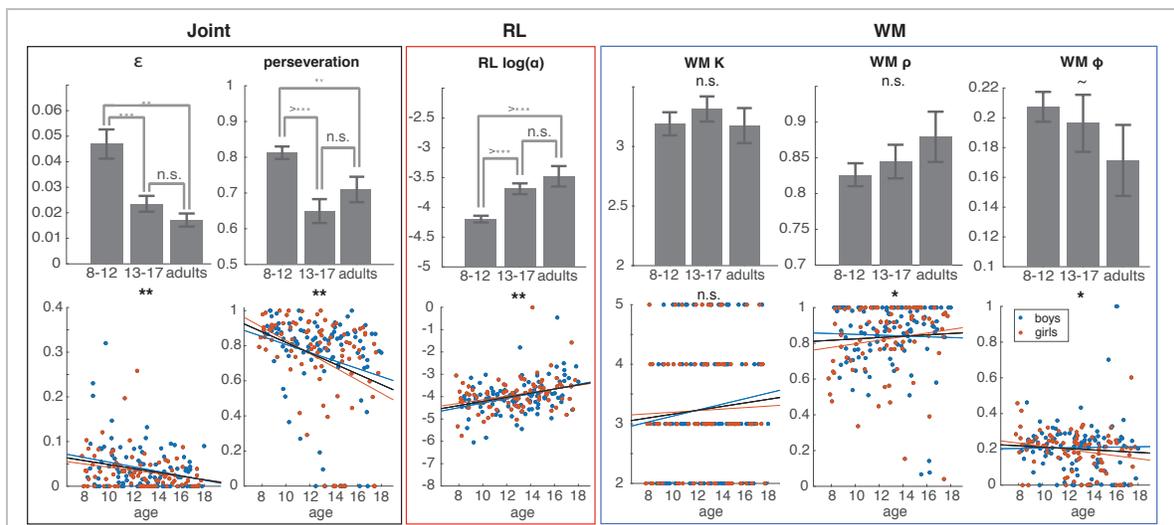
WM, RL, or both drove learning improvement with age. While a decrease in set size effect could hint at a WM effect, it is equally possible that worse performance in high set sizes in younger children could be due to worse RL (as hinted by the logistic regression results), which can support learning when WM is unavailable (e.g. at high set sizes). To clarify these findings, we next sought to model individual participants' behavior with a mixture model capturing both RL and WM contributions to learning (see Methods).

**Modeling.** Model comparison favored the RLWM model over other candidate models in all age groups (Figure 3ab; see methods). The exceedance probability in favor of the RLWM model was 1 in all groups (Rosa, Bestmann, Harrison, & Penny, 2010). RLWM was also the best model out of 6 candidate models for 83 out of 100 8 to 12-year-olds, 58 out of 59 13 to 17-year-olds, and 25 out of 54 25 to 30-year-olds (Figure 3a). Model simulations with fit parameters reproduced subject behavior, as well as differences between age groups (Figure 3c, supplementary figure S1).

We first investigated two noise parameters (Figure 4: Joint, epsilon and perseveration) and as expected, there was an effect of age group on the decision noise parameter epsilon (Kruskall-Wallis  $p = 0.025$ ). Post-hoc comparison revealed that this was driven by the separation in behavior between children 8-12 and the older age groups (Rank-sum test 8 to 12-year-olds vs. 13 to 17-year-olds:  $p = 0.0003$ ; 8 to 12-year-olds vs. 25 to 30-year-olds:  $p = 0.002$ ; 13 to 17-year-olds vs. 25 to 30-year-olds:  $p = 0.94$ ). There was a negative relationship between age and decision noise in the 8 to 17-year-old sample

( $\rho = -0.25$ ,  $p = 0.0005$ ), in boys ( $\rho = -0.22$ ,  $p = 0.024$ ), and in girls ( $\rho = -0.3$ ,  $p = 0.005$ ), showing that decisions were less noisy in older participants. There was also an effect of age group on the perseveration parameter (Kruskal-Wallis  $p = 0.0003$ ). Post-hoc comparisons also revealed that the difference between 8 to 12-year-olds and the other age groups drove the main effect of group (Rank-sum test 8 to 12-year-olds vs. 13 to 17-year-olds:  $p < 10e-4$ ; 8 to 12-year-olds vs. 25 to 30-year-olds:  $p = 0.001$ ; 13 to 17-year-olds vs. 25 to 30-year-olds:  $p = 0.67$ ). In the 8-17-year-old sample, there was also a negative relationship between age and perseveration ( $\rho = -0.3$ ,  $p = 1.9e-07$ ). This relationship was present separately in both boys ( $\rho = -0.3$ ,  $p = 0.002$ ) and girls ( $\rho = -0.44$ ,  $p = 1.9e-05$ ),

We next investigated the relationship between age and RL learning rate  $\alpha$ . There was a robust effect of age group on the RL learning rate parameter (Figure 4: RL; Kruskal-Wallis  $p = 0.0002$ ). Additional comparison between groups showed that 8 to 12-year-olds had a lower average learning rate than 13 to 17-year-olds ( $p < 10e-4$ ) and 25 to 30-year-olds ( $p < 10e-4$ ). 13 to 17-year-olds did not differ from 25 to 30-year-olds ( $p = 0.3$ ). Individual learning rates showed a significant upward trend with age 8-17 ( $\rho = 0.31$ ,  $p = 2e-05$ ), which was significant separately in both boys ( $\rho = 0.32$ ,  $p = 0.001$ ) and girls ( $\rho = 0.29$ ,  $p = 0.007$ ).



**Figure 4. Effects of age on RLWM model parameters.** There was an effect of both age group and age in years on the perseveration and  $\epsilon$  decision noise parameters, whereby 8-12-year-olds had noisier behavior, and integrated negative feedback less than 13-17-year-olds and 25-30-year-olds. There were robust effects of age on the RL learning rate

parameter  $\alpha$ . There was no effect of coarse age groups or continuous effect of age within the 8-17 sample on the WM capacity parameter, and weak effects on WM weight  $\rho$  and decay  $\phi$ . Error bars on bar graphs are standard error of the mean for each age group.  $\sim$  indicates marginal significance at the  $p < 0.1$  level, \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$ , n.s. stands for not significant.

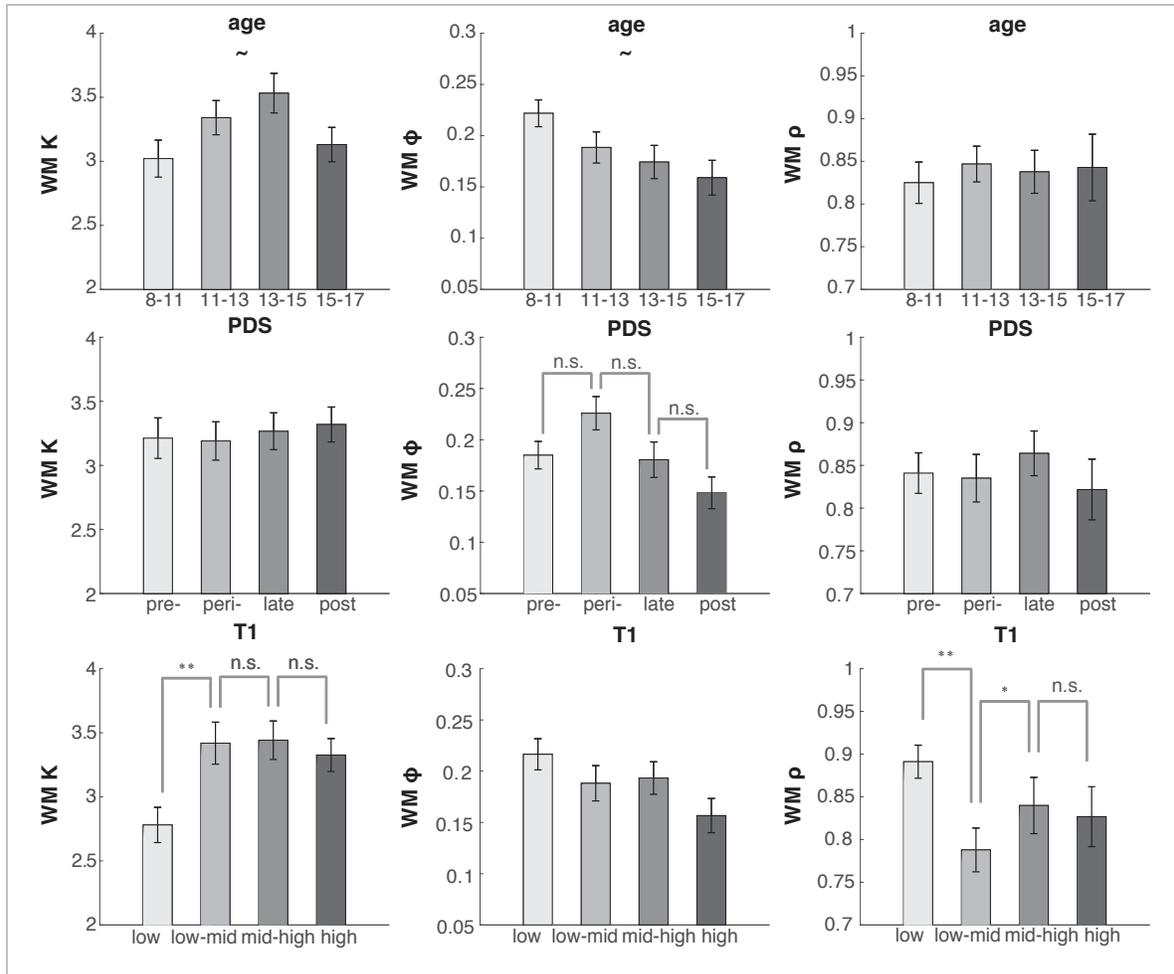
Finally, we investigated WM parameters. Surprisingly, we found no significant differences across our coarse age groups in WM capacity (Kruskal-Wallis test,  $p = 0.83$ ), (Figure 4: WM left) (although see PDS and testosterone group results below). When we focused on individual subjects aged 8 to 17, we found no monotonic relationship between WM capacity and age (Figure 4: WM; Spearman  $\rho = 0.09$ ,  $p = 0.23$ ). We found no differences across coarse age groups in WM weight (Figure 4: WM;  $p = 0.25$ ), but a marginal effect on WM decay (Figure 4;  $p = 0.08$ ). These effects were stronger when investigating the continuous effect of age within non-adults: There was a small positive relationship between age and WM weight (Spearman  $\rho = 0.15$ ,  $p = 0.04$ ; Boys:  $\rho(101) = 0.06$ ,  $p = 0.58$ ; Girls:  $\rho(86) = 0.25$ ,  $p = 0.02$ ). There was also a negative relationship between age and WM decay (Spearman  $\rho = -0.17$ ,  $p = 0.02$ ) that was inconsistent across genders (Boys:  $\rho = -0.08$ ,  $p = 0.42$ ; Girls:  $\rho = -0.28$ ,  $p = 0.01$ ). This continuous effect was mostly driven by the youngest participants, with more WM decay and less WM weight.

Children of the same age can differ in their stage of pubertal maturation with considerable individual variability as well as sex differences in the timing of pubertal onset. Pubertal onset and later pubertal milestones may also produce non-monotonic changes over time such as a ‘U’ or ‘inverted U’ shape curve ((Piekarski, Boivin, et al., 2017; Piekarski, Johnson, et al., 2017)). Gonadal hormones, such as testosterone and its metabolites may also be playing an activational role in cortical or basal ganglia function on the day of testing ((Delevich, Thomas, et al., 2019)). Therefore, to explore developmental changes in learning with a finer resolution relevant to puberty we divided the 8-17 year old sample into 4 evenly divided bins first by age (<10.5, 10.6-12.8, 12.9-14.8, and >14.9 years old) then by pubertal development scale (PDS) (roughly pre-, early, mid-, and late/post-pubertal) and then by salivary testosterone (low, low-mid, mid-high, and high levels within gender) at time of test (T1) (see methods, Supplemental material). Cross-sectional data are quite limiting in differentiating age and pubertal effects, however

we carefully explored the data and the effects of age, PDS, and testosterone concentration.

Division by finely graded age bins showed that noise and perseveration decreased with age (Supplementary Figures S4, S5; Noise: chi-squared = 15.69,  $p = 0.0013$ ; pers: chi-squared = 24.1,  $p = 2e-05$ ) and RL learning rate increased significantly with age (Supplementary Figure S6; chi-squared = 19.3,  $p = 0.0002$ ) with no inverted U patterns. This general pattern was present in both boys (Supplementary Figures S4, S5, and S6; RL learning rate: chi-squared = 12.1,  $p = 0.007$ ; noise: chi-squared = 6.7,  $p = 0.08$ ; pers: chi-squared = 9.1,  $p = 0.028$ ) and girls (RL learning rate: chi-squared = 8.1,  $p = 0.044$ ; noise: chi-squared = 10.13,  $p = 0.018$ ; pers: chi-squared = 26.2,  $p = 0.001$ ). We observed similar results when non-adult participants were divided into bins based on PDS or testosterone measures (see supplementary table S2). Age, PDS and mean salivary testosterone were all also highly correlated (Supplementary Figure S2), therefore it was not clear if puberty onset played a significant role in RL related changes or not.

We next examined the relationship between finely graded age groups age, PDS and salivary testosterone measures and WM parameters. Here the greater number of groups enabled observation of ‘U’ or ‘inverted U shapes.’ Patterns of this type were apparent when subjects were grouped by PDS or T1. We had predicted potential effects of puberty onset on WM due to our previous work on effects of gonadal hormones on inhibitory neurotransmission in the PFC of rodent models.



(Piekarski, Boivin, et al., 2017; Piekarski, Johnson, et al., 2017). Subjects that were grouped by PDS did show significant differences in WM decay that was stronger in boys than girls (see Figure 5, Supplemental Figure 8; all: chi-squared = 10.02,  $p = 0.018$ ; boys: chi-squared = 15.09,  $p = 0.0017$ ; girls: chi-squared = 5.51,  $p = 0.14$ ). Changes were strongest at puberty onset. We also uncovered a significant effect of testosterone at time of test (T1) on WM capacity (see Figure 5, Supplemental Figure 9; chi-squared = 13.15,  $p = 0.0043$ ). This effect was marginal in boys alone (chi-squared = 6.61,  $p = 0.085$ ) and

significant in girls alone (chi-squared = 7.86,  $p = 0.049$ ). When subjects were grouped by age (in 4 bins age 8-17), without regard to PDS or T, no significant group differences in WM parameters were found (though there was a marginal inverse U-shaped effect on WM capacity, and decreasing effect on decay; Figure 5, Supplemental Figures 8 and 9).

These results support our first prediction that changes in both RL and WM processes separately drive learning changes in subjects aged 8 to 30. However, we were surprised to find that changes in RL drive changes in learning more than changes in WM, in opposition to our second prediction. Finally, we support our third prediction that WM processes would differ in groups separated by pubertal development and testosterone.

## Discussion

In this study, we examined developmental changes in the working memory (WM) and reinforcement learning (RL) processes that contribute to simple stimulus-action association learning. While many developmental models emphasize late prefrontal cortex maturation and gains in WM in late adolescence ((Huizinga et al., 2006; Larsen & Luna, 2018)), previous studies have also shown changes in development in reinforcement learning tasks (Palmineri et al., 2016; Potter et al., 2017; Van Den Bos et al., 2012). However, these studies were not able to disentangle WM and RL contributions to learning. Our task and computational model were designed to address how WM and RL jointly contribute to learning in a simple task. ((Collins, 2018; Collins, Ciullo, Frank, & Badre, 2017; Collins & Frank, 2012; Collins et al., 2017)). The task and model are also notable because they reveal that WM is recruited in even simple reward learning tasks that are often assumed to include only model-free RL processes.

Here we used the RLWM task and computational methods to measure development of RL- and WM-based processes in children and teenagers aged 8 to 17 and adults aged 25 to 30. We found that 8 to 12-year-olds, 13 to 17-year-olds, and 25 to 30-year-olds all recruited both RL and WM in parallel for learning. 13 to 17-year-olds' performance closely approximated 25 to 30-year-olds', while participants aged 8 to 12 learned more slowly and reached a lower asymptotic accuracy. This was partially due to younger subjects' noisier behavior, which may be indicative of a higher attentional lapse rate or

tendency to explore different options, and to a lack of integration of negative feedback. However, we also related lower performance in younger subjects to a weaker contribution of RL to learning and to a smaller extent, to weaker WM maintenance and use.

Indeed, our behavioral results and computational model did not support our second hypothesis that WM changes are the main drivers of learning improvement from 8 to 17 years of age. Instead our data supported only our first hypothesis that RL-based learning develops from age 8 to 17, while some aspects of WM processes also did, though less strongly. Counter-intuitively, while 8 to 12-year-olds' behavior was more sensitive to the learning load, we found that this was characteristic of a weaker RL system unable to make up for WM's limitations in high set size conditions, more so than reflecting weaker capacity-limited WM. In addition to age effects on the decision noise and perseveration parameters, we found that the model's RL learning rate ( $\alpha$ ) parameter increased with age; by contrast there were no significant effects of age on WM capacity. Decision noise and perseveration decreased with age while the RL learning rate steadily increased throughout adolescence and reached adult levels in 13 to 17-year-olds. Given the learning environment (a deterministic task with an unchanging reward structure), an optimal strategy would be to learn quickly from rewards (i.e. have a higher learning rate). This allows learners to reach asymptotic accuracy faster as long as the environment is stable. Thus 13 to 17-year-olds and 25 to 30-year-olds learned stimulus-action associations faster than 8 to 12-year-olds, which benefitted their overall task performance. We did observe small effects of age on WM weight and WM decay, reflecting increased use and stability of WM with age. Also, pubertal development and salivary testosterone concentrations showed relationships with WM decay and capacity, respectively (at puberty onset).

We attempted to disentangle the influences of age and puberty on behavior despite the limitations of cross-sectional data and of our indirect measures of puberty, which is a complex multi-component process. Pubertal development was also highly correlated with age in our sample, however there were late and early developing individuals within our sample. When we examined noise, perseveration, and RL learning rate and divided subjects by pubertal development (PDS) or testosterone (T)

concentrations, we largely observed the same patterns that were observed when using age. However, when we examined WM parameters in groups divided by PDS and T, we uncovered ‘inverted U’ shaped trajectories and found that WM decay changed significantly after pubertal onset in both boys and girls. WM capacity was also lowest in the group with the lowest testosterone concentrations at time of test. Thus pubertal status and gonadal hormone levels may also play a role in the development of WM decay and WM capacity, supporting our third hypothesis that there would be puberty-related changes in WM.

This study supports our first hypothesis but does not support our second hypothesis because we expected to observe strong age-related effects on WM processes. Instead, we observed the strongest age-related effects on RL learning rate. The strongest developmental effects on WM parameters were those of testosterone and pubertal development.

Adolescence is widely associated with the development of higher cognitive resources and the use of increasingly complex learning strategies ((Crone et al., 2006; Decker et al., 2016; Potter et al., 2017; Selmeczy et al., 2018)). These shifts in learning are mirrored by the development of prefrontal cortex, which does not reach functional or structural maturation until late in the second decade or even middle of the third decade of life ((Bunge & Gabrieli, 2002; Casey, Jones, & Hare, 2008)). Recent work proposes that adolescence as a life stage extends until age 25 ((Sawyer et al., 2018)). Furthermore, the use of WM in complex tasks has been shown to improve during development into late adolescence ((Crone et al., 2006; Geier et al., 2009; Huizinga et al., 2006; Luna, 2009; McAuley & White, 2011)). Therefore, it is somewhat surprising that we detected no age-related changes in WM capacity in subjects aged 8 to 17 despite our large sample. This is partially mitigated by PDS and T-related findings that show changes with the transition into puberty and the fact that other aspects of working memory (decay and weight) did show significant if weak effects of age. To better integrate this work into the established literature on the developmental science of learning, it may be helpful to dissociate the use of WM to maintain information from its use to manipulate information. The use of WM to manipulate information may develop over adolescence ((Crone et al., 2006)), while the use of WM to maintain information perhaps develops earlier. Recent results with a simple

WM assay, a verbal span task, found age groups aged 9 to 25 performed equally well ((Potter et al., 2017)) and thus found no developmental effects on WM. The lack of age effect on the WM capacity in our study may reflect the fact that WM is used to maintain stimulus-action associations, rather than to manipulate information. However, in this task, information was never directly given to participants for them to hold – instead, they needed to integrate three temporally separate pieces of information (the stimulus observed, the choice made, and the feedback received), and determine the relevant information to maintain in WM. This, we would argue, constitutes an advanced mental manipulation in WM.

While dual systems models of development have often highlighted that subcortical areas mature before prefrontal cortex, there is also an important literature showing changes in striatal function through adolescence. Adolescents in the mid teen years have been shown more sensitive to and motivated by rewards than pre-pubertal children and post-pubertal adults ((Casey et al., 2008; Braams et al., 2015; Davidow, Foerde, Galván, & Shohamy, 2016; Gulley et al., 2018; Somerville, Jones, & Casey, 2010)). This increased motivation is reflected both in BOLD activation of the reward-responsive ventral striatum (Braams et al., 2015; A. Galvan et al., 2006b) and in increased recruitment of areas of the frontoparietal cognitive control network, when necessary ((Somerville & Casey, 2010;)). This could partially account for our results, explaining why 13 to 17-year-olds use RL more efficiently than 8 to 12-year-olds. However, such a relationship would predict differences between 13 to 17-year-olds and 25 to 30-year-olds, as seen in learning tasks that manipulate reinforcement more directly ((Palminteri et al., 2016)). We did not observe any difference between subjects aged 13 to 17 and adults in both behavior and in modeling; this might be due to a comparatively weak reinforcement manipulation, though it was sufficient to reveal strong differences between 8 to 12-year-olds and 13 to 17-year-olds.

An additional way to interpret our results might be to consider the role of episodic memory in our task. Recent research showed that a component of learning from reinforcement can be accounted for by episodic memory sampling ((Bornstein & Daw, 2012; Bornstein & Norman, 2017; Myers et al., 2002)). In fact, previous versions of this task found hippocampal contributions to learning ((Collins et al., 2017)). Thus, it is

possible that some behavior captured by the RL component of the model actually included contributions of the episodic memory system, not just the reward learning systems. Some facets of episodic memory, like recognition memory, have been shown to develop by age 8, much earlier than prefrontal cortex-dependent executive function ((Ghetti & Bunge, 2012)). However, recent work shows that the interactions between hippocampus and prefrontal cortex – crucial for efficient memory search and retrieval – continue developing into adolescence ((Murty, Calabro, & Luna, 2016; Selmeczy et al., 2018)). Further, the development of both systems allows for the integration of relevant past experiences with goal-directed attention and can influence choice in the reward-driven BG ((Murty et al., 2016)). This developmental time course might account for some of the changes attributed to RL learning rate in our results. It will be an important topic for future research to better characterize episodic memory contributions to learning, in parallel to working memory and reinforcement learning's contributions.

**Conclusions.** Using a sample of 241 participants from 8 to 30 years old, we aimed to characterize the distinct developmental trajectories of multiple cognitive systems that contribute to learning in parallel, even in very simple situations: the reinforcement learning and working memory systems. Performance in a deterministic reward stimulus-action learning task which manipulated cognitive load revealed broad differences between subjects aged 8 to 12 and 13 to 17, whose performance was adult-like. While 8 to 12-year-olds were noisier in all conditions, their learning suffered more from a load increase than 13 to 17-year-olds' and 25 to 30-year-olds'. In all participants, RL learning compensated for WM limitations under higher load. Computational modeling revealed that the stronger effect of load in 8 to 12-year-olds was less due to weaker time-limited, capacity-limited WM than to weaker RL compensation. This result was surprising based on the established literature on the late anatomical development of the prefrontal cortex. This work highlights the importance of carefully accounting for multiple systems' contributions to learning when assessing group and individual differences and suggests that the development of reinforcement learning processes plays a major role in changes in learning during adolescent development. We hope these findings can fruitfully inform educational methods and intervention work. Future research in the science of learning

should aim to develop experimental paradigms and computational models that more precisely dissociate the use of reinforcement learning, working memory, and episodic memory throughout development.

## Bibliography

- Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron*, *66*(2), 315–326.  
<https://doi.org/10.1016/j.neuron.2010.03.025>
- Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience*, *35*(7), 1011–1023. <https://doi.org/10.1111/j.1460-9568.2011.07920.x>
- Bornstein, A. M., & Norman, K. A. (2017). Reinstated episodic context guides sampling-based decisions for reward. *Nature Neuroscience*, *20*(7), 997–1003.  
<https://doi.org/10.1038/nn.4573>
- Braams, B. R., van Duijvenvoorde, A. C. K., Peper, J. S., & Crone, E. A. (2015). Longitudinal Changes in Adolescent Risk-Taking: A Comprehensive Study of Neural Responses to Rewards, Pubertal Development, and Risk-Taking Behavior. *Journal of Neuroscience*, *35*(18), 7226–7238.  
<https://doi.org/10.1523/jneurosci.4764-14.2015>
- Bunge Silvia A., M., D. N., E., T. M., J., V. C., & Gabrieli D.E., J. (2002). Immature Frontal Lobe Contributions to Cognitive Control in Children: Evidence from fMRI. *Neuron*, *33*, 301–311. [https://doi.org/10.1016/S0896-6273\(01\)00583-9](https://doi.org/10.1016/S0896-6273(01)00583-9)
- Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The adolescent brain. *Annals of the New York Academy of Sciences*, *1124*, 111–126. <https://doi.org/10.1196/annals.1440.010>
- Cohen, J. R., Asarnow, R. F., Sabb, F. W., Bilder, R. M., Bookheimer, S. Y., Knowlton, B. J., & Poldrack, R. A. (2010). A unique adolescent response to reward prediction errors. *Nature Neuroscience*, *13*(6), 669–671. <https://doi.org/10.1038/nn.2558>
- Collins, A. G. E. (2018). The tortoise and the hare: Interactions between reinforcement learning and working memory. *Journal of Cognitive Neuroscience*, *30*, 1422–1432.
- Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working Memory Load Strengthens Reward Prediction Errors. *The Journal of Neuroscience*, *37*(16), 4332–

4342. <https://doi.org/10.1523/jneurosci.2700-16.2017>
- Collins, A. G. E., & 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.  
<https://doi.org/10.1111/j.1460-9568.2011.07980.x>
- Collins, A. G. E., & Frank, M. J. (2014). *NIH Public Access*. *120*(1), 190–229.  
<https://doi.org/10.1037/a0030852>.Cognitive
- Collins, A. G. E., & Frank, M. J. (2018). Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proceedings of the National Academy of Sciences*, 201720963.  
<https://doi.org/10.1073/pnas.1720963115>
- Crone, E. A., Wendelken, C., van Leijenhorst, L., Donohue, S., & Bunge, S. A. (2006). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences*, *103*(24), 9315–9320.  
<https://doi.org/10.1073/pnas.0510088103>
- Curtis, C. E., & D’Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*(9), 415–423.  
[https://doi.org/10.1016/S1364-6613\(03\)00197-9](https://doi.org/10.1016/S1364-6613(03)00197-9)
- D’Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annu Rev Psychol.*, *66*(66), 115–142. <https://doi.org/10.1146/annurev-psych-010814-015031>.THE
- Davidow, J. Y., Foerde, K., Galván, A., & Shohamy, D. (2016). An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. *Neuron*, *92*(1), 93–99. <https://doi.org/10.1016/j.neuron.2016.08.031>
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans’ choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. <https://doi.org/10.1016/j.neuron.2011.02.027>
- Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From Creatures of Habit to Goal-Directed Learners. *Psychological Science*, *27*(6), 848–858.  
<https://doi.org/10.1177/0956797616639301>
- Delevich, K., Piekarski, D., & Wilbrecht, L. (2019). Neuroscience: Sex Hormones at

- Work in the Neocortex. *Current Biology*, 29(4), R122–R125.  
<https://doi.org/10.1016/j.cub.2019.01.013>
- Delevich, K., Thomas, A. W., & Wilbrecht, L. (2019). Adolescence and “Late Blooming” Synapses of the Prefrontal Cortex. *Cold Spring Harbor Symposia on Quantitative Biology*, LXXXIII, 037507. <https://doi.org/10.1101/sqb.2018.83.037507>
- Galvan, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., & Casey, B. J. (2006a). Earlier Development of the Accumbens Relative to Orbitofrontal Cortex Might Underlie Risk-Taking Behavior in Adolescents. *Journal of Neuroscience*, 26(25), 6885–6892. <https://doi.org/10.1523/JNEUROSCI.1062-06.2006>
- Galvan, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., & Casey, B. J. (2006b). Earlier Development of the Accumbens Relative to Orbitofrontal Cortex Might Underlie Risk-Taking Behavior in Adolescents. *Journal of Neuroscience*, 26(25), 6885–6892. <https://doi.org/10.1523/JNEUROSCI.1062-06.2006>
- Galvan, Adriana, Delevich, K., & Wilbrecht, L. (2019). Cortico-Striatal Circuits and Changes in Reward, Learning, and Decision Making in Adolescence. In Gazzaniga (Ed.), *The Cognitive Neurosciences, 6th edition*.
- Geier, C. F., Garver, K., Terwilliger, R., & Luna, B. (2009). Development of Working Memory Maintenance. *Journal of Neurophysiology*, 101(1), 84–99.
- Ghetti, S., & Bunge, S. A. (2012). Neural changes underlying the development of episodic memory during middle childhood. *Developmental Cognitive Neuroscience*, 2(4), 381–395. <https://doi.org/10.1016/j.dcn.2012.05.002>
- Giedd, J. N., Blumenthal, J., & Jeffries, N. O. et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neurosci* 1999; 2: 861–63. *Nature Neuroscience*, 2, 861–863.
- Gold, J. M., Waltz, J. A., Frank, M. J., Albrecht, M. A., & Collins, A. G. E. (2017). Interactions Among Working Memory, Reinforcement Learning, and Effort in Value-Based Choice: A New Paradigm and Selective Deficits in Schizophrenia. *Biological Psychiatry*, 82(6), 431–439.  
<https://doi.org/10.1016/j.biopsych.2017.05.017>
- Gulley, J. M., Willing, J., Paul, M. J., Flores, C., Walker, D. M., & Bell, M. R. (2018). Adolescence and Reward: Making Sense of Neural and Behavioral Changes Amid

- the Chaos. *The Journal of Neuroscience*, 37(45), 10855–10866.  
<https://doi.org/10.1523/jneurosci.1834-17.2017>
- Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Toward an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. *Modelling Natural Action Selection*, (April), 239–263.  
<https://doi.org/10.1017/CBO9780511731525.016>
- Herting, M. M., Gautam, P., Spielberg, J. M., Kan, E., Dahl, R. E., & Sowell, E. R. (2014). The role of testosterone and estradiol in brain volume changes across adolescence: A longitudinal structural MRI study. *Human Brain Mapping*, 35(11), 5633–5645. <https://doi.org/10.1002/hbm.22575>
- Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, 44(11), 2017–2036.  
<https://doi.org/10.1016/j.neuropsychologia.2006.01.010>
- Juraska, J., & Willing, J. (2017). Pubertal onset as a critical transition for neural development and cognition. *Brain Research*, 1654(B), 87–94.  
<https://doi.org/10.1016/j.brainres.2016.04.012>
- Larsen, B., & Luna, B. (2018). Adolescence as a neurobiological critical period for the development of higher-order cognition. *Neuroscience and Biobehavioral Reviews*, 94(June), 179–195. <https://doi.org/10.1016/j.neubiorev.2018.09.005>
- Luna, B. (2009). Developmental Changes in Cognitive Control through Adolescence. *Adv Child Dev Behav*, 37, 233–278.  
<https://doi.org/10.1146/annurev.biochem.78.082307.091526>.Regulation
- McAuley, T., & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of Experimental Child Psychology*, 108(3), 453–468.  
<https://doi.org/10.1016/j.jecp.2010.08.009>
- Miller, E. K., & Cohen, J. D. (2001). *AN INTEGRATIVE THEORY OF PREFRONTAL CORTEX FUNCTION*. 167–202.
- Murty, V. P., Calabro, F., & Luna, B. (2016). The role of experience in adolescent cognitive development: Integration of executive, memory, and mesolimbic systems.

- Neuroscience and Biobehavioral Reviews*, 70, 46–58.  
<https://doi.org/10.1016/j.neubiorev.2016.07.034>
- Myers, C., Shohamy, D., Clark, J., Gluck, M. A., Paré-Blagoev, E. J., Creso Moyano, J., & Poldrack, R. A. (2002). Interactive memory systems in the human brain. *Nature*, 414(6863), 546–550. <https://doi.org/10.1038/35107080>
- Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., ... Ward, G. (2018). Benchmarks of models of short-term and working memory. *Psychological Bulletin*. <https://doi.org/10.5962/bhl.title.7369>
- Palminteri, S., Kilford, E. J., Coricelli, G., & Blakemore, S. J. (2016). The Computational Development of Reinforcement Learning during Adolescence. *PLoS Computational Biology*, 12(6), 1–25. <https://doi.org/10.1371/journal.pcbi.1004953>
- Peper, J. S., Hulshoff Pol, H. E., Crone, E. A., & van Honk, J. (2011). Sex steroids and brain structure in pubertal boys and girls: A mini-review of neuroimaging studies. *Neuroscience*, 191, 28–37. <https://doi.org/10.1016/j.neuroscience.2011.02.014>
- Petersen, C., & Crockett, L. (1988). <Puberty Scale.Pdf>. (2).  
<https://doi.org/10.1103/PhysRevLett.78.1396>
- Piekarski, D. J., Boivin, J. R., & Wilbrecht, L. (2017). Ovarian Hormones Organize the Maturation of Inhibitory Neurotransmission in the Frontal Cortex at Puberty Onset in Female Mice. *Current Biology*, 27(12), 1735-1745.e3.  
<https://doi.org/10.1016/j.cub.2017.05.027>
- Piekarski, D. J., Johnson, C. M., Boivin, J. R., Thomas, A. W., Lin, W. C., Delevich, K., ... Wilbrecht, L. (2017). Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? *Brain Research*, 1654, 123–144.  
<https://doi.org/10.1016/j.brainres.2016.08.042>
- Potter, T. C. S., Bryce, N. V., & Hartley, C. A. (2017). Cognitive components underpinning the development of model-based learning. *Developmental Cognitive Neuroscience*, 25, 272–280. <https://doi.org/10.1016/j.dcn.2016.10.005>
- Rosa, M. J., Bestmann, S., Harrison, L., & Penny, W. (2010). Bayesian model selection maps for group studies. *NeuroImage*, 49(1), 217–224.  
<https://doi.org/10.1016/j.neuroimage.2009.08.051>
- Sawyer, S. M., Azzopardi, P. S., Wickremarathne, M., & Patton, G. C. (2018). The age of

- adolescence. *The Lancet Child & Adolescent Health*, 2(3), 223–228.
- Selmecky, D., Fandakova, Y., Grimm, K. J., Bunge, S. A., & Ghetti, S. (2018). Longitudinal trajectories of hippocampal and prefrontal contributions to episodic retrieval: Effects of age and puberty. *Developmental Cognitive Neuroscience*, (June), 100599. <https://doi.org/10.1016/j.dcn.2018.10.003>
- Somerville, L. H., & Casey, B. J. (2010). Developmental neurobiology of cognitive control and motivational systems. *Current Opinion in Neurobiology*, 20(2), 236–241. <https://doi.org/10.1016/j.conb.2010.01.006>
- Somerville, L. H., Jones, R. M., & Casey, B. J. (2010). A time of change: Behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain and Cognition*, 72(1), 124–133. <https://doi.org/10.1016/j.bandc.2009.07.003>
- Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, 9(2), 69–74. <https://doi.org/10.1016/j.tics.2004.12.005>
- Sutton, R. S., & Barto, A. G. (2017). *Reinforcement Learning: An introduction* (2nd ed.). 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. <https://doi.org/10.1093/cercor/bhr198>
- Wilson, R., & Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *Submitted*.
- Wimmer, G. E., Braun, E. K., Daw, N. D., & Shohamy, D. (2014). Episodic Memory Encoding Interferes with Reward Learning and Decreases Striatal Prediction Errors. *Journal of Neuroscience*, 34(45), 14901–14912. <https://doi.org/10.1523/jneurosci.0204-14.2014>

## Supplementary Materials

### Supplemental Methods

**Pubertal development scores.** Subjects' pubertal development scores (PDS) were assessed using the Petersen Development Scale ((Petersen & Crockett, 1988)). Subjects were asked to assess their pubertal development with 5 questions. Items assessed growth in height, growth of body hair, skin changes, and development of secondary sex characteristics. Subjects were asked to rate with one of four choices how development had progressed from “not started yet” to “has completed.” Ratings were assigned values of 1 - 4. All answered questions were averaged to produce a PDS. The lowest possible PDS is a 1 while the highest possible is a 4.

**Age, PDS, and Testosterone bins.** To better understand the independent effects of age, pubertal development, and mean testosterone on model parameters, subjects aged 8 to 17 were binned into one of four groups for each measure. Age bins were created by splitting all subjects aged 8 to 17 into four quartiles (see Table 1). The age range (mean) of each bin is 8.06 to 10.54 (9.32), 10.62 to 12.85 (11.88), 12.92 to 14.86 (13.89), and 14.88 to 17.99 (16.31) years old.

PDS bins were created in a similar fashion as the age bins. Subjects were split into gender-specific quartiles based on the range of PDS scores of all subjects of their gender, then combined into one bin. Since the PDS is derived from a questionnaire with 5 questions, there were many subjects with the same PDS. As such, the bins are not exactly equal in size (see Table 1). The mean PDS in each bin is 1.19, 1.86, 2.80, and 3.59.

Testosterone bins were also created with a quartile procedure. Within subjects under 18, subjects were split into gender-specific quartiles, one based on mean testosterone from 1 to 3 samples per subject and one based on testosterone concentration from sample 1, the

in-lab sample taken in the middle of the experimental session. Subjects of the same gender-specific quartile (i.e. the males and females with the lowest levels of testosterone within their gender) were then binned together. The mean (std) testosterone levels of all mean testosterone bins were 18.04 (5.78), 33.65 (6.3), 58.78 (17.19), and 102.80 (34.83) picograms/mL. The mean (std) testosterone levels of the sample 1 testosterone bins were 14.68 (4.79), 26.04 (3.38), 43.15 (9.19), and 100.49 (51.43) picograms/mL.

**Table 1:** Subject N's in each of four bins of age, PDS, mean testosterone, and testosterone from sample 1:

	<b>low</b>	<b>mid</b>	<b>mid-high</b>	<b>high</b>
Age (M/F)	47 (26/21)	47 (26/21)	47 (26/21)	46 (24/22)
Mean (min-max)	9.32 (8.06 – 10.54)	11.88 (10.62 – 12.85)	13.89 (12.92 – 14.86)	16.31 (14.88 – 17.99)
PDS (M/F)	47 (26/21)	42 (23/19)	45 (25/20)	47 (25/22)
Mean	1.19	1.86	2.80	3.59
Mean T (M/F)	42 (23/19)	44 (24/20)	44 (24/20)	47 (25/22)
Mean (std)	18.04 (5.78)	33.65 (6.3)	58.78 (17.19)	102.80 (34.83)
T1 (M/F)	41 (22/19)	43 (23/20)	43 (23/20)	46 (24/22)
Mean (std)	14.68 (4.79)	26.04 (3.38)	43.15 (9.19)	100.49 (51.43)

## Supplemental Results

**Pubertal development.** To assess the specific effect of pubertal development on learning, we further tested the relationship between behavioral and modeling measures and pubertal development scale (PDS) and salivary testosterone. Age and PDS were highly correlated in adolescents aged 8 to 17 (Fig. S1,  $\rho = 0.92$ ,  $p = 0$ ), and in boys and girls separately (both  $\rho$ 's  $> 0.75$ ,  $p$ 's  $< 1.e-19$ ). Age and mean testosterone were also highly correlated in the whole sample ( $\rho = 0.52$ ,  $p = 3e-17$ ), in boys and men ( $\rho = 0.77$ ,  $p = 2e-25$ ), and in girls and women, though less so ( $\rho = 0.31$ ,  $p = 9e-04$ ). Testosterone and PDS were also highly correlated in all adolescents aged 8 to 17 and in both genders

separately (Fig. S1, Adolescents:  $\rho = 0.47$ ,  $p = 8e-22$ ; Boys:  $\rho = 0.79$ ,  $p = 3. e-21$ ; Girls:  $\rho = 0.5$ ,  $p = 2.1e-06$ ). It is therefore difficult to disentangle effects of pubertal status from the effects of age in adolescents aged 8 to 18. Indeed, all analyses we conducted that included PDS or testosterone while controlling for age did not show effects of puberty. However, this could be due to the high correlation and does illuminate whether or not pubertal changes are potential causal variables. As such, we ran group and continuous analyses with each measure not accounting for the other measures and compared effects across measures.

The results of those analyses are presented in the main text. In summary, PDS results strongly support age results. There are also some PDS results in WM weight and decay. In contrast, testosterone results support age results for RL learning rate, noise and perseveration parameters, but relate to individual differences in WM capacity, in contrast to the effects on WM weight and decay for age and puberty measures.

## Group Results

**Table 2:** Kruskal-wallis tests on age groups, PDS, mean testosterone (T), and testosterone from sample 1 (T1) groups (all subjects aged 8 to 17):

	Age group chi-squared	Age group p	PDS group chi-squared	PDS group p	T group chi-squared	T group p	T1 group chi-squared	T1 group p
RL alpha	19.29	<b>0.0002</b>	10.9	<b>0.0123</b>	9.89	<b>0.042</b>	7.33	<b>0.062</b>
Epsilon	15.69	<b>0.0013</b>	13.86	<b>0.0031</b>	16.13	<b>0.002</b>	8.96	<b>0.029</b>
Perseveration	24.13	<b>2.34e-05</b>	9.86	<b>0.0198</b>	15.79	<b>0.003</b>	17.09	<b>0.0007</b>
WM decay	6.25	<b>0.1</b>	10.02	<b>0.018</b>	6.8	0.14	5.2	<b>0.16</b>
WM weight	4.87	0.18	1.18	0.75	0.63	0.96	9.32	<b>0.025</b>
WM capacity	7.56	<b>0.056</b>	0.62	0.89	15.8	<b>0.003</b>	13.15	<b>0.004</b>

A series of post-hoc non-parametric T-tests revealed that the lowest testosterone group exhibited smaller WM capacity than all other testosterone groups ( $p$ 's < 0.01). The low-mid, mid-high, and high testosterone groups exhibited no difference in WM capacity.

**Table 3:** Kruskal-wallis tests on age, PDS, mean T, and T1 groups (male subjects only):

<b>Boys</b>	Age group chi-squared	Age group p	PDS group chi-squared	PDS group p	T group chi-squared	T group p	T1 group chi-squared	T1 group p
RL alpha	12.09	<b>0.007</b>	6.96	<b>0.07</b>	12.21	0.016	9.08	<b>0.028</b>
Epsilon	6.71	<b>0.082</b>	6.26	<b>0.09</b>	7.7	0.1	2.03	0.56
Perseveration	9.14	<b>0.028</b>	2.37	0.49	12.28	<b>0.015</b>	10.01	<b>0.019</b>
WM decay	1.66	0.65	15.09	<b>0.0017</b>	5.9	0.21	3.37	0.33
WM weight	1.27	0.73	1.49	0.68	2.83	0.59	3.75	0.28
WM capacity	7.19	<b>0.066</b>	2.3	0.51	6.06	0.19	6.61	<b>0.0855</b>

**Table 4:** Kruskal-wallis tests on age, PDS, mean T, and T1 groups (female subjects only):

<b>Girls</b>	Age group chi-squared	Age group p	PDS group chi-squared	PDS group p	T group chi-squared	T group p	T1 group chi-squared	T1 group p
RL alpha	8.09	<b>0.044</b>	4.32	0.23	1.89	0.75	0.59	0.89
Epsilon	10.13	<b>0.018</b>	13.06	<b>0.0045</b>	11.61	<b>0.021</b>	10.24	<b>0.017</b>
Perseveration	26.17	<b>0.001</b>	10.44	<b>0.0152</b>	4.47	0.34	7.22	<b>0.065</b>
WM decay	5.6	0.13	5.51	0.13	4.57	0.33	5.37	0.14

WM weight	5.1	0.16	2.5	0.47	0.83	0.93	5.61	0.13
WM capacity	1.6	0.66	0.86	0.83	12.07	0.017	7.86	<b>0.049</b>

### Continuous Results within 8-17 year-olds

**Table 5:** Spearman correlations of summary behavioral measures, and PDS, mean T, and T1:

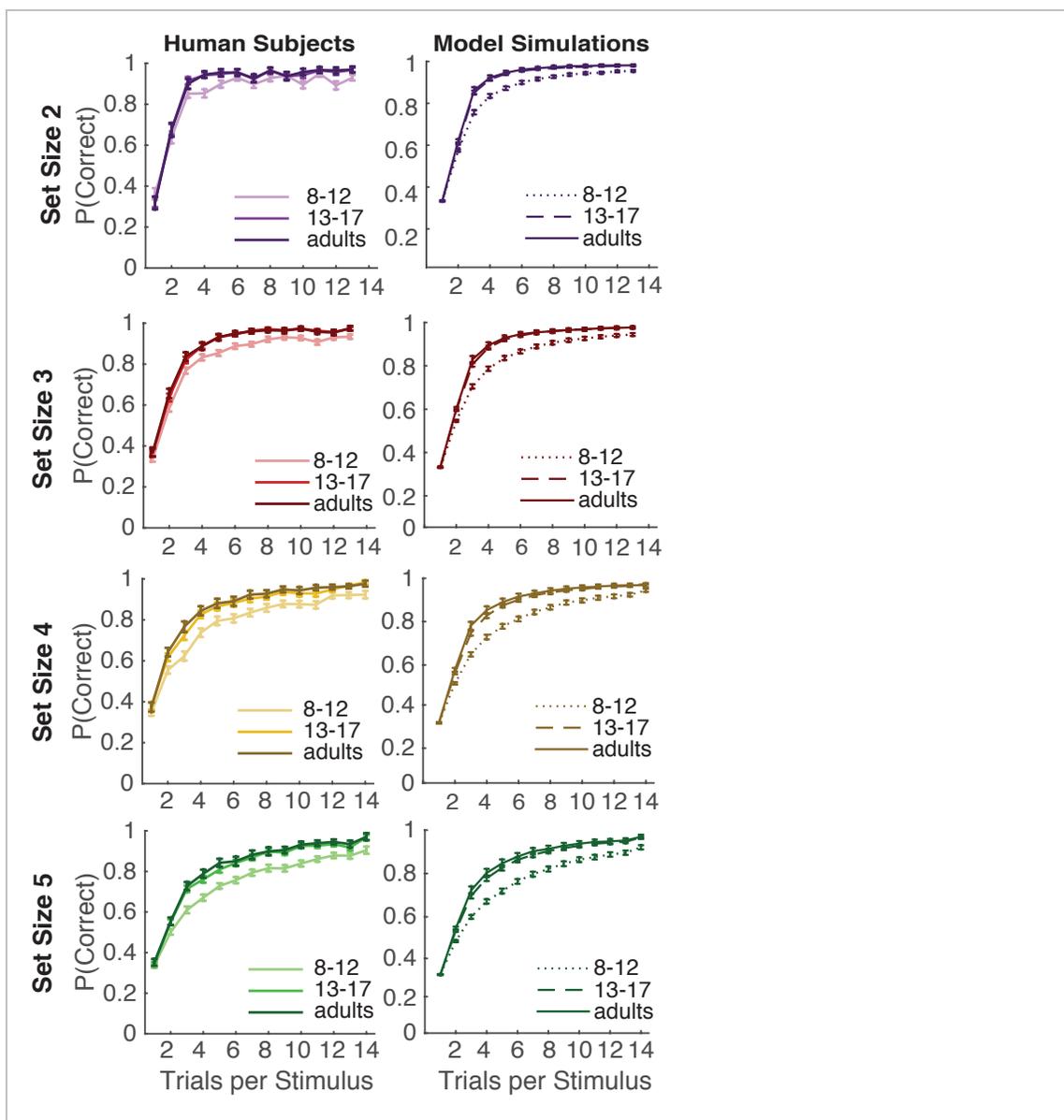
<b>Girls</b>	PDS rho	PDS p	Mean T rho	Mean T p	T1 rho	T1 p
Overall accuracy	0.351	<b>0.001</b>	0.285	<b>0.0025</b>	0.343	<b>0.0025</b>
Ns slope	-0.207	<b>0.059</b>	-0.1901	<b>0.045</b>	-0.148	0.12
<b>Boys</b>						
Overall accuracy	0.324	<b>0.0011</b>	0.48	<b>2.03e-08</b>	0.464	<b>1.39e-07</b>
Ns slope	-0.09	0.354	-0.154	0.091	-0.173	<b>0.062</b>

**Table 6:** Spearman correlations of model parameters, and age, PDS, and mean T in all subjects aged 8 to 17:

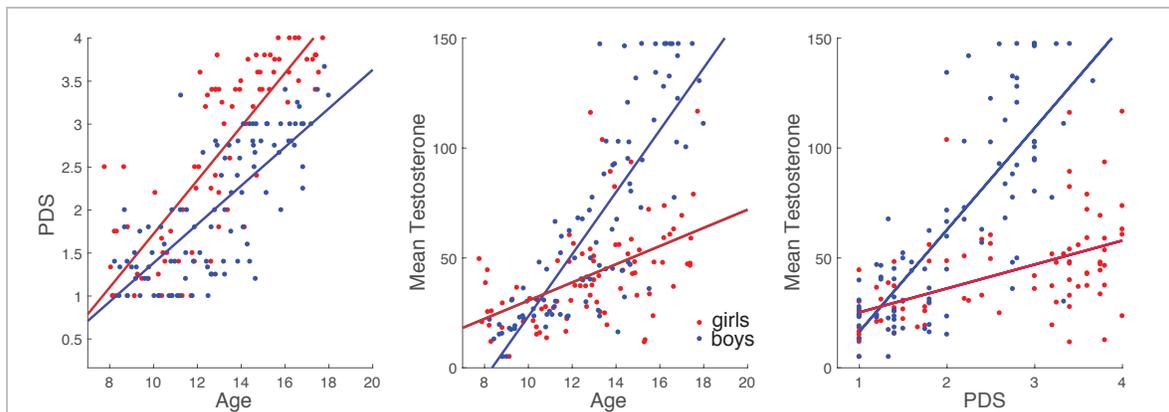
Parameter	Age rho	Age p	PDS rho	PDS p	Mean T rho	Mean T p
RL alpha	0.31	<b>1.96e-05</b>	0.21893	<b>0.002</b>	0.21	<b>0.003</b>
Epsilon	-0.25	<b>0.0005</b>	-0.269	<b>0.0002</b>	-0.24243	<b>0.001</b>
Perseveration	-0.37	<b>2e-07</b>	-0.233	<b>0.0015</b>	-0.26	<b>0.0004</b>
WM decay	-0.16	<b>0.019</b>	-0.11	0.12	-0.09	0.21
WM weight	0.15	<b>0.034</b>	0.03	0.59	0.0103	0.89
WM capacity	0.08	0.23	0.05	0.49	0.17	<b>0.021</b>

**Table 7:** Spearman correlations of model parameters, and age, PDS, and mean T in girls and boys aged 8 to 17:

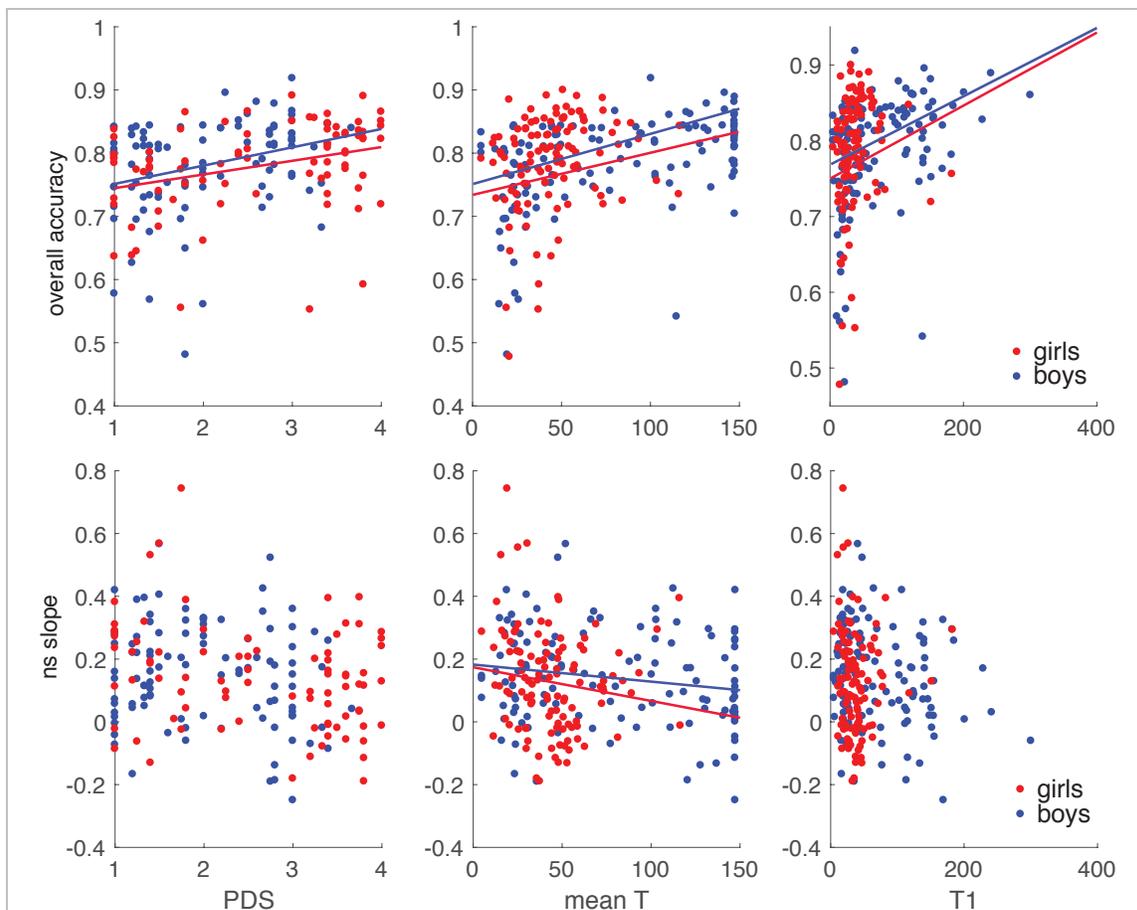
<b>GIRLS (n = 87)</b>	<b>Age rho</b>	<b>Age p</b>	<b>PDS rho</b>	<b>PDS p</b>	<b>Mean T rho</b>	<b>Mean T p</b>
RL alpha	0.286	<b>0.007</b>	0.20	<b>0.0677</b>	0.11	0.3
Epsilon	-0.3	<b>0.005</b>	-0.33	<b>0.002</b>	-0.332	<b>0.002</b>
Perseveration	-0.44	<b>1.963e-05</b>	-0.29	<b>0.006</b>	-0.2031	0.066
WM decay	-0.27	<b>0.007</b>	-0.2005	<b>0.069</b>	-0.159	0.15
WM weight	0.24	<b>0.022</b>	0.119	0.28	-0.07	0.49
WM capacity	0.02	0.81	0.017	0.8745	0.11	0.3
<b>BOYS (n = 100)</b>	<b>Age rho</b>	<b>Age p</b>	<b>PDS rho</b>	<b>PDS p</b>	<b>Mean T rho</b>	<b>Mean T p</b>
RL alpha	0.319	<b>0.001</b>	0.211	<b>0.036</b>	0.32	<b>0.001</b>
Epsilon	-0.223	<b>0.024</b>	-0.228	<b>0.023</b>	-0.19	<b>0.061</b>
Perseveration	-0.304	<b>0.002</b>	-0.112	0.27	-0.32	<b>0.001</b>
WM decay	-0.081	0.42	0.005	0.96	-0.089	0.389
WM weight	0.06	0.48	-0.04	0.67	0.071	0.48993
WM capacity	0.14	0.14	0.089	0.38	0.12	0.24



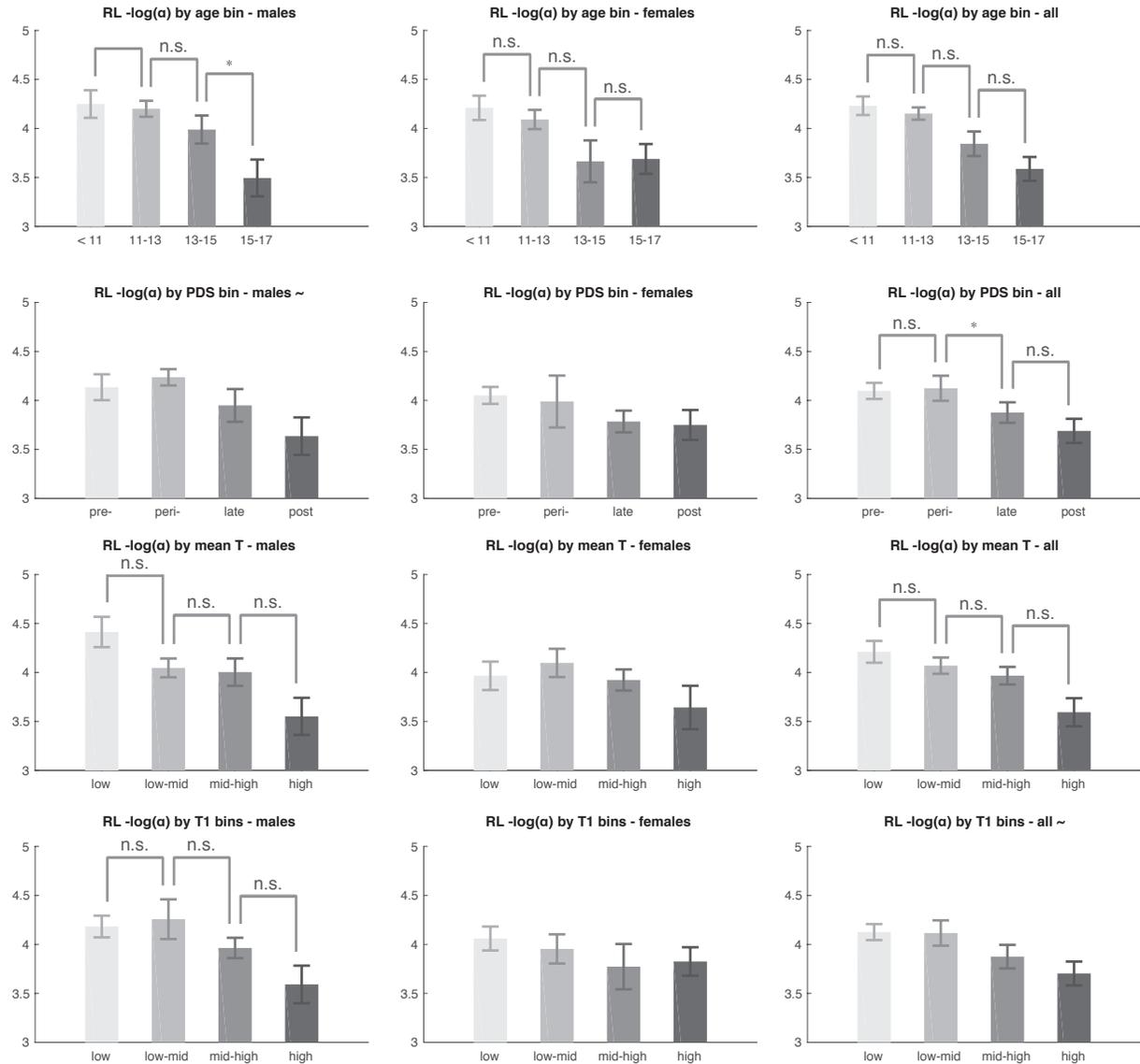
**Supplemental Figure 1. Model validation.** Proportion correct per stimulus iteration in set sizes 2 – 5 for both human subjects and the RLWM model (see main Figure 3). Model simulation reproduces human behavior at all set sizes. The model can reproduce both age and set size effects when behavior is simulated with parameters fit to each subject individually.



**Supplemental Figure 2. Relationships between puberty measures.** In the full sample of adolescents aged 8 to 17, age, pubertal development score (PDS), and mean salivary testosterone are highly correlated. Each of these relationships also stands in boys and girls separately, though boys' mean testosterone is more related to both age and PDS than girls'.

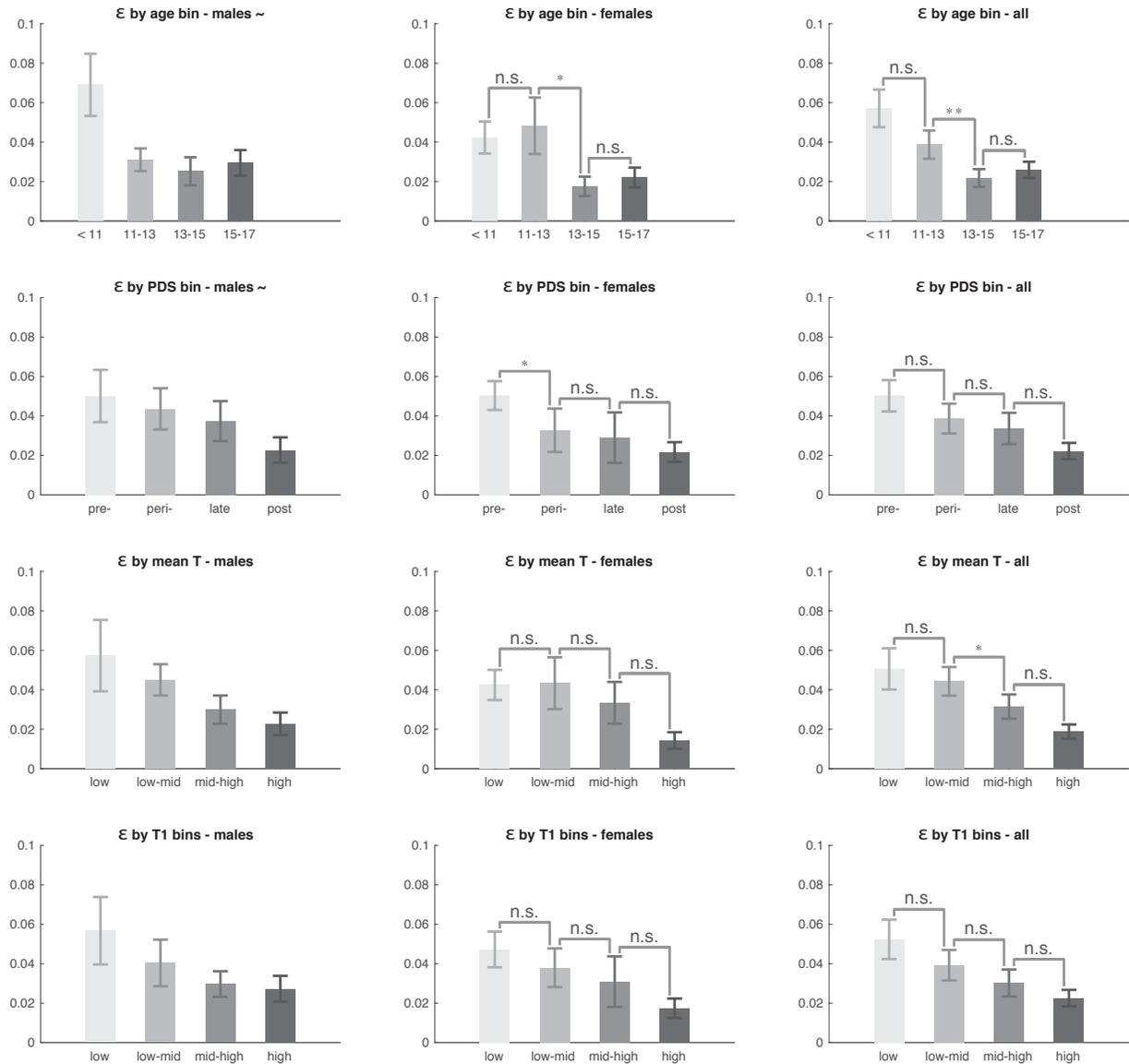


**Supplemental figure 3. Behavioral summary measures by puberty measures.** The relationships of pubertal development score (PDS), mean testosterone concentration across 1 to 3 saliva samples (mean T), and testosterone concentration in the in-lab sample (T1). Non-parametric Spearman correlations were run in girls and boys aged 8 to 17 separately. In girls, there was a positive effect on overall accuracy of PDS ( $\rho = 0.35$ ,  $p = 0.0012$ ), mean T ( $\rho = 0.29$ ,  $p = 0.0025$ ), and T1 ( $\rho = 0.34$ ,  $p = 0.0025$ ). The same was true in boys (PDS: ( $\rho = 0.32$ ,  $p = 0.0011$ , Mean T:  $\rho = 0.48$ ,  $p = 2.03e-08$ , T1:  $\rho = 0.46$ ,  $p = 1.39e-07$ ). The only effect of puberty on the set size effect (ns slope) in girls was that of mean T ( $\rho = -0.19$ ,  $p = 0.046$ ), though this effect was only marginal in boys ( $\rho = -0.15$ ,  $p = 0.091$ ). There was also a marginal relationship of PDS and ns slope in girls ( $\rho = -0.207$ ,  $p = 0.059$ ) and of T1 and ns slope in boys ( $\rho = -0.17$ ,  $p = 0.062$ ).

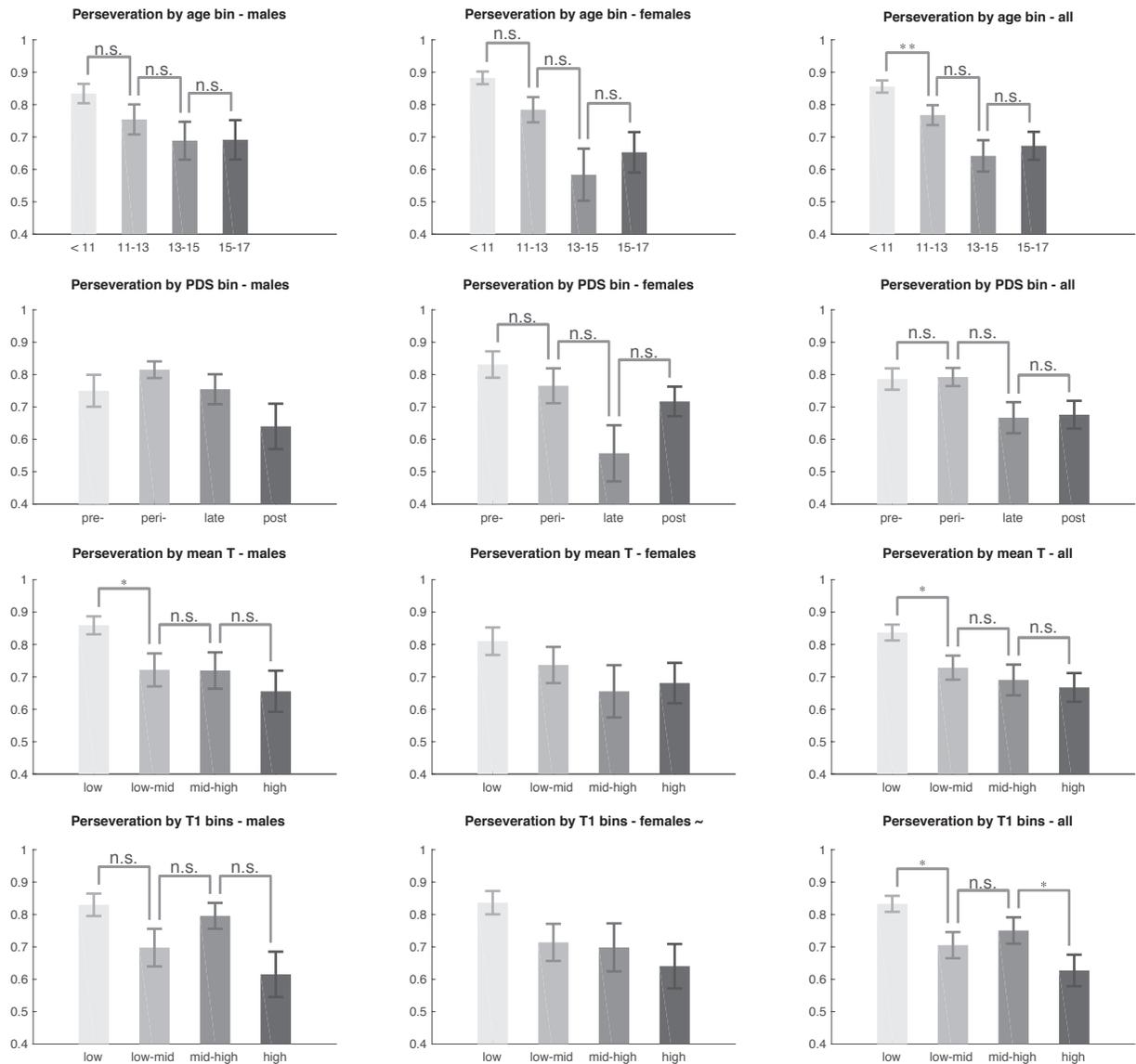


**Supplemental Figure 4. RL  $\alpha$  (learning rate) by age, PDS, and testosterone (T) bins.**

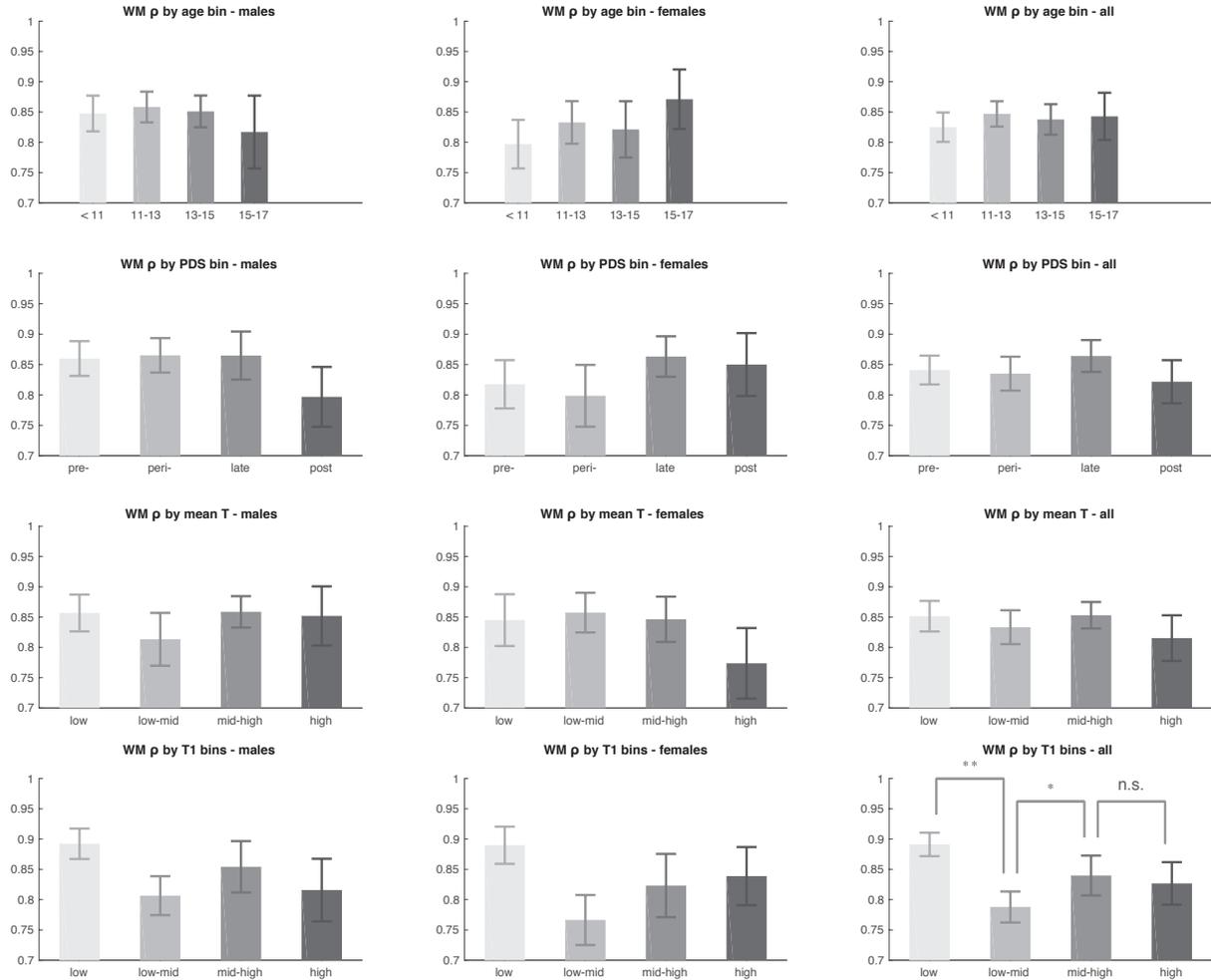
Male and female subjects (8-17 years old) were binned into four groups using three different measures: age, PDS, or mean testosterone. Mean RL log( $\alpha$ ) was then plotted for each of the four groups, in males alone, females alone, and the combined sample. For Kruskal-Wallis group p-values at or below a threshold of  $p = 0.05$ , post-hoc comparisons were run and their results plotted. For pairwise comparisons n.s. indicates not significant at  $p < 0.05$  threshold, \*  $p < 0.05$ , \*\*  $p < 0.01$  for pairwise comparisons; For Kruskal-Wallis tests ~ indicates marginal significance at  $p < 0.1$  level. Error bars represent the standard error of the mean. In both male and female subjects, we found significant group effects and PDS effects of age on learning rate.



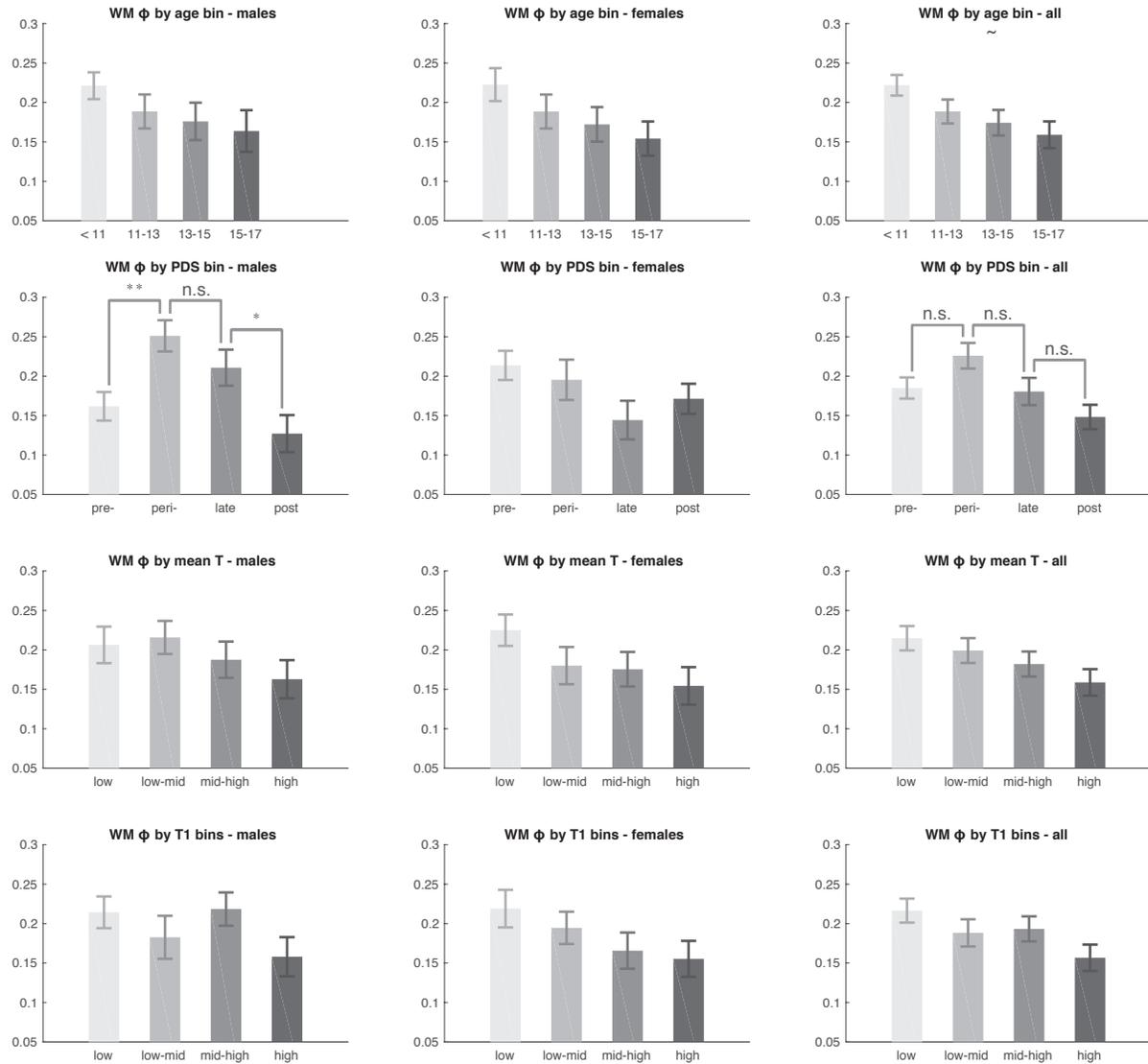
**Supplemental Figure 5.  $\epsilon$  (noise) by age, PDS, mean testosterone (T), and sample 1 testosterone (T1) bins.** Male and female subjects (8-17 years old) were binned into four groups using three different measures: age, PDS, or mean testosterone. Mean decision noise was then plotted for each of the four groups, in males alone, females alone, and the combined sample. For Kruskal-Wallis group p-values at or below a threshold of  $p = 0.05$ , post-hoc comparisons were run and their results plotted. For pairwise comparisons, n.s. indicates not significant at  $p < 0.05$  threshold, \*  $p < 0.05$ , \*\*  $p < 0.01$  for pairwise comparisons; For Kruskal-Wallis tests ~ indicates marginal significance at  $p < 0.1$  level. Error bars represent the standard error of the mean. We observed significant age, PDS, and T effects on epsilon in the entire sample, but only marginal effects in boys when separated from girls.



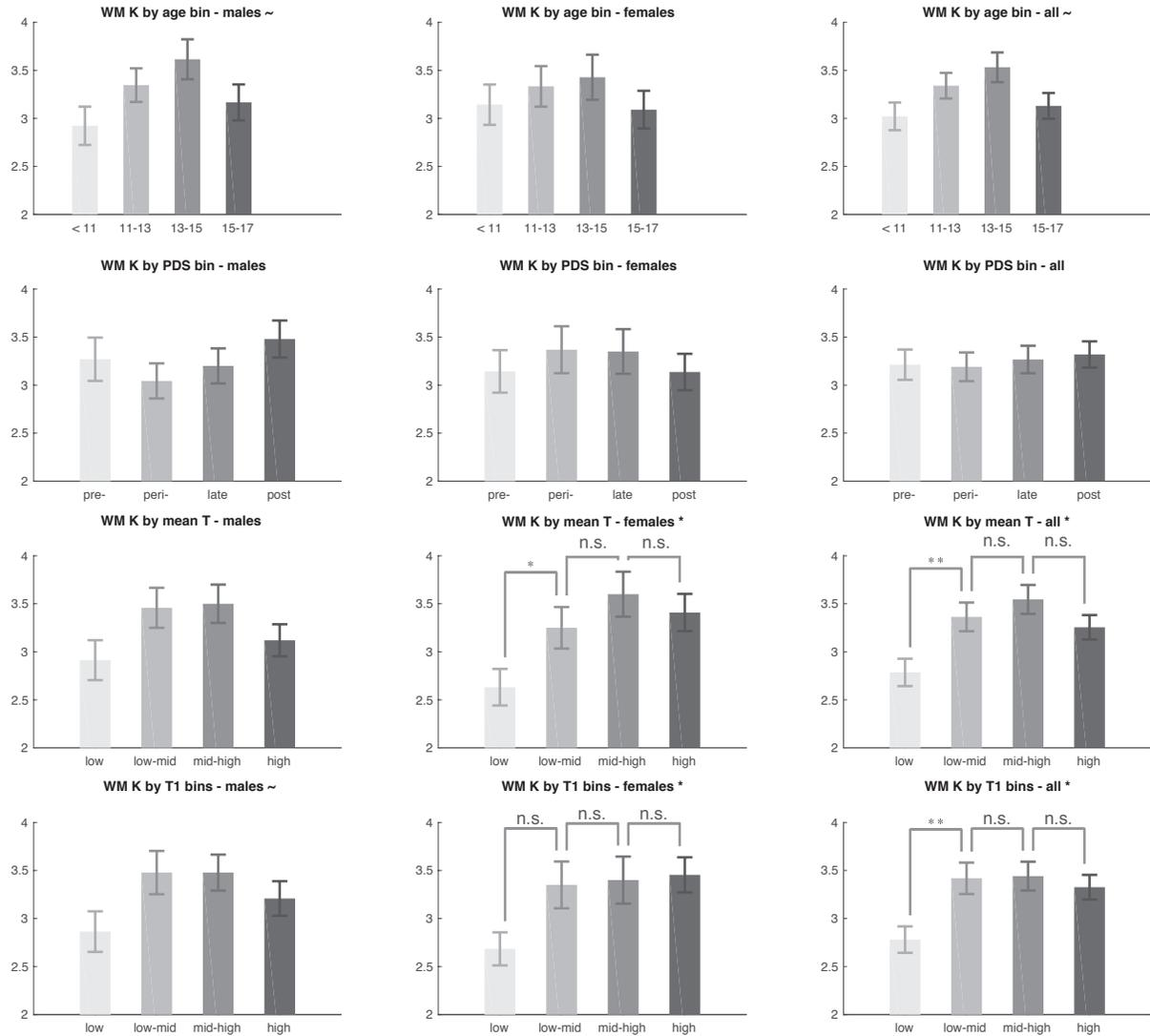
**Supplemental Figure 6. Perseveration by age, PDS, mean testosterone (T), and sample 1 testosterone (T1) bins.** Male and female subjects (8-17 years old) were binned into four groups using three different measures: age, PDS, or mean testosterone. Mean perseveration parameter was then plotted for each of the four groups, in males alone, females alone, and the combined sample. For Kruskal-Wallis group p-values at or below a threshold of  $p = 0.05$ , post-hoc comparisons were run and their results plotted. For pairwise comparisons n.s. indicates not significant at  $p < 0.05$  threshold, \*  $p < 0.05$ , \*\*  $p < 0.01$  for pairwise comparisons; For Kruskal-Wallis tests ~ indicates marginal significance at  $p < 0.1$  level. Error bars represent the standard error of the mean. Perseveration decreased with age in a graded way across boys, girls, and the entire sample of 8 to 17-year-olds.



**Supplemental Figure 7. WM  $\rho$  (weight) by age, PDS, mean testosterone (T), and sample 1 testosterone (T1) bins.** Male and female subjects (8-17 years old) were binned into four groups using three different measures: age, PDS, mean testosterone, or testosterone from sample 1 (T1). Mean WM weight was then plotted for each of the four groups, in males alone, females alone, and the combined sample. For Kruskal-Wallis group p-values at or below a threshold of  $p = 0.05$ , post-hoc comparisons were run and their results plotted. For pairwise comparisons n.s. indicates not significant at  $p < 0.05$  threshold, \*  $p < 0.05$ , \*\*  $p < 0.01$  for pairwise comparisons; For Kruskal-Wallis tests ~ indicates marginal significance at  $p < 0.1$  level. Error bars represent the standard error of the mean. We observed an effect of T1, activational testosterone, in the whole sample of 8 to 17-year-old subjects on WM use. There were no other development-related effects on WM use.



**Supplemental Figure 8. WM  $\phi$  (decay) by age, PDS, mean testosterone (T), and sample 1 testosterone (T1) bins.** Male and female subjects (8-17 years old) were binned into four groups using three different measures: age, PDS, mean testosterone, or testosterone from sample 1 (T1). Mean WM decay was then plotted for each of the four groups, in males alone, females alone, and the combined sample. For Kruskal-Wallis group p-values at or below a threshold of  $p = 0.05$ , post-hoc comparisons were run and their results plotted. For pairwise comparisons n.s. indicates not significant at  $p < 0.05$  threshold, \*  $p < 0.05$ , \*\*  $p < 0.01$  for pairwise comparisons; For Kruskal-Wallis tests ~ indicates marginal significance at  $p < 0.1$  level. Error bars represent the standard error of the mean. We observed an inverse U-shaped effect of pubertal development score (PDS) on WM decay in males and on the combined sample. Post-hoc PDS bin comparisons revealed that pre- and post-pubertal males appear to drive the overall effect of PDS on decay. The effect of age only reached marginal significance in the combined sample.



**Supplemental Figure 9. WM K (capacity) by age, PDS, mean testosterone (T), and sample 1 testosterone (T1) bins.** Male and female subjects (8-17 years old) were binned into four groups using three different measures: age, PDS, mean testosterone, or testosterone from sample 1 (T1). Mean WM capacity was then plotted for each of the four groups, in males alone, females alone, and the combined sample. For Kruskal-Wallis group p-values at or below a threshold of  $p = 0.05$ , post-hoc comparisons were run and their results plotted. For pairwise comparisons n.s. indicates not significant at  $p < 0.05$  threshold, \*  $p < 0.05$ , \*\*  $p < 0.01$  for pairwise comparisons; For Kruskal-Wallis tests ~ indicates marginal significance at  $p < 0.1$  level. Error bars represent the standard error of the mean. There were robust effects of both mean testosterone and activational sample 1 testosterone on mean WM capacity. Closer examination of those effects revealed that subjects with the lowest level of sample 1 and mean T exhibited lower WM capacity.