

Cognitive components underpinning the development of model-based learning



Tracey C.S. Potter^{a,b,1}, Nessa V. Bryce^{a,b,1}, Catherine A. Hartley^{a,b,*}

^a Sackler Institute for Developmental Psychobiology, Weill Cornell Medicine, New York, NY 10065, United States

^b New York University, Department of Psychology, New York, NY 10003, United States

ARTICLE INFO

Article history:

Received 1 July 2016

Received in revised form 9 August 2016

Accepted 20 October 2016

Available online 29 October 2016

Keywords:

Model-based

Reinforcement learning

Fluid reasoning

Statistical learning

ABSTRACT

Reinforcement learning theory distinguishes “model-free” learning, which fosters reflexive repetition of previously rewarded actions, from “model-based” learning, which recruits a mental model of the environment to flexibly select goal-directed actions. Whereas model-free learning is evident across development, recruitment of model-based learning appears to increase with age. However, the cognitive processes underlying the development of model-based learning remain poorly characterized. Here, we examined whether age-related differences in cognitive processes underlying the construction and flexible recruitment of mental models predict developmental increases in model-based choice. In a cohort of participants aged 9–25, we examined whether the abilities to infer sequential regularities in the environment (“statistical learning”), maintain information in an active state (“working memory”) and integrate distant concepts to solve problems (“fluid reasoning”) predicted age-related improvements in model-based choice. We found that age-related improvements in statistical learning performance did not mediate the relationship between age and model-based choice. Ceiling performance on our working memory assay prevented examination of its contribution to model-based learning. However, age-related improvements in fluid reasoning statistically mediated the developmental increase in the recruitment of a model-based strategy. These findings suggest that gradual development of fluid reasoning may be a critical component process underlying the emergence of model-based learning.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Individuals can recruit a variety of evaluative strategies to make everyday decisions. Reinforcement learning theory distinguishes two such strategies: model-based and model-free learning (Daw et al., 2005, 2011; Glascher et al., 2010). Model-based learning requires the construction of a cognitive model of potential actions and their consequences, which can be consulted to determine the best way to pursue a current goal. Such learning supports flexible behavior in novel situations and can readily take into account changes in the environment. By contrast, model-free learning simply estimates the value of reflexively repeating an action based on whether it previously led to good or bad outcomes, without representing the specific outcomes themselves. While model-free learning is computationally efficient, it cannot rapidly adjust to

changes in the value of an outcome or changes in contingency between an action and outcome.

Many decisions or actions can be evaluated in a model-based or a model-free manner. Effective behavioral control often involves striking a context-dependent balance between these deliberative versus automatic strategies. Recent research suggests that while model-free learning is consistently employed across developmental stages, recruitment of model-based learning tends to increase with age (Decker et al., 2016). Across diverse decision-making contexts or tasks, younger individuals exhibit patterns of behavior that reflect greater reliance on a model-free strategy, whereas older individuals rely more on model-based learning (Decker et al., 2016; Klosssek et al., 2008; Piaget, 1954; Zelazo et al., 1996). The developmental timepoint at which one typically shifts toward employing a model-based strategy may depend on both the intrinsic complexity of the task at hand, as well as the maturity of the myriad cognitive processes required for the formation and recruitment of a mental model of that task.

To make goal-directed decisions, individuals must be able to anticipate likely events, consider the consequences of their potential actions, and evaluate the most efficient means to obtain a

* Corresponding author at: New York University, 6 Washington Place, Suite 888, United States.

E-mail address: cah369@nyu.edu (C.A. Hartley).

¹ These authors contributed equally to this work.

desired outcome. The ability to recognize which events tend to follow each other in sequence or covary with high probability is often referred to as statistical learning (Turk-Browne et al., 2005). Simple forms of statistical learning are present in infants and children (Amso and Davidow, 2012; Fiser and Aslin, 2002), demonstrating that individuals can build cognitive models of environmental statistics from early on in development. However, in other tasks, statistical learning performance has been observed to improve with age (Schlichting et al., 2016), suggesting that learning of more complex sequential structures may emerge later in development. More accurate representations of the statistical structure of a task may facilitate model-based choice. However, whether increased recruitment of model-based learning with age might reflect developmental improvements in statistical learning remains an open question.

Developmental changes in the reliance on model-based learning might also reflect an increasing capacity to recruit learned cognitive models to guide decisions. Working memory, the ability to maintain mental representations in an active state despite interference, is a key component of model recruitment (D'Esposito and Postle, 2015). Introducing working memory load during decision-making reduces adults' use of a model-based strategy (Otto et al., 2013a), and high working memory capacity buffers individuals from stress-induced impairment of model-based learning (Otto et al., 2013b). Another important process potentially underlying successful model recruitment is fluid reasoning, the capacity to flexibly integrate independent goal-relevant associations across domains. Fluid reasoning involves the reorganization, transformation, and extrapolation of learned conceptual relationships in order to solve novel problems (Cattell, 1987; McArdle et al., 2002). Both working memory and fluid reasoning have been shown to increase from early childhood into young adulthood (Ferrer et al., 2009; Fry and Hale, 1996), suggesting that either of these processes, or their integrated function, may foster increased recruitment of model-based choice.

Building upon a previous finding that model-based reinforcement learning increased with age from childhood into adulthood (Decker et al., 2016), in this study, we sought to characterize the cognitive underpinnings of this developmental trajectory. Given previous observations of age-related changes in statistical learning, working memory, and fluid reasoning, we examined the contributions of these putative component processes to the development of model-based choice in a sequential reinforcement-learning task. We found that fluid reasoning, but not statistical learning, mediated the relationship between age and model-based choice. Ceiling performance on our working memory assay prevented examination of its contribution to model-based learning. Collectively, these findings suggest that the protracted development of fluid reasoning ability may be a critical process underpinning the gradual emergence of model-based learning.

2. Methods

2.1. Participants

22 children (aged 9–12), 23 adolescents (13–17), and 24 adults (18–25) took part in this study. All participants, and parents of minors, provided written informed consent according to the procedures of the Weill Cornell Medical College Institutional Review Board and received monetary compensation for participation. Subjects completed a sequential reinforcement-learning task while undergoing a functional MRI scan. Neuroimaging data are not analyzed or reported here. Subjects also completed a statistical learning task, and two subtests of the Wechsler Abbreviated Scale of Intelligence (WASI, matrix-reasoning and vocabulary

sections). Subjects who missed more than 15 trials (10% of trials) during the reinforcement-learning task were excluded from analysis, leaving 19 children (13 females, 10.5 ± 1.1 years), 22 adolescents (12 females, 14.7 ± 1.5 years) and 23 adults (14 females, 21.6 ± 2.1 years) in the final sample. Of these participants, statistical learning task data for 1 child was not acquired due to a computer malfunction, 1 adolescent and 2 adults did not complete the WASI matrix-reasoning subtest, and 1 adolescent and 2 adults did not complete the WASI vocabulary subtest. A subset of participants (14 children, 17 adolescents, 18 adults) also completed the listening recall subtest of the Automated Working Memory Assessment.

2.2. Reinforcement-learning task

The two-stage sequential reinforcement-learning task was adapted for developmental populations by Decker et al. (2016) from a task designed by Daw et al. (2011) to dissociate model-based and model-free evaluative strategies (Fig. 1A). In this paradigm, participants were tasked with collecting space treasure, and were told they would be paid a monetary bonus based on the amount of space treasure that they found. At the first stage of each trial, participants selected one of two spaceships (“first-stage choice”) that would make a probabilistic transition to a red or purple planet. Each spaceship transitioned to one planet more frequently than the other (70% of trials versus 30%). These “common” and “rare” transition probabilities did not change during the task. Once at a planet, participants then selected one of two aliens to ask for space treasure (“second-stage choice”). Each alien provided treasure according to a slowly drifting probability of reward. Subjects had three seconds to make a choice at each stage.

The task was designed to dissociate use of a model-based strategy, in which individuals recruit a mental model of the task's probabilistic state transition structure, from use of a model-free strategy, which requires only cached estimates of the past rewards associated with preceding first-stage actions.

All participants played a 50-trial tutorial to become familiar with the structure of the task before completing the 150-trial task in the scanner; the tutorial and full versions of the task had different colored stimuli but the same task structure and rules. During the tutorial, participants were instructed that each spaceship usually went to a specific planet, but had to learn the transitions and probabilities themselves from the task. All subjects, regardless of performance, received a fixed bonus payment at the end of the scan.

Using a previously described analytical approach (Daw et al., 2011), we fit a hybrid reinforcement-learning model to participants' choice data. The hybrid model allows participants' choices to reflect a weighted average of both model-free and model-based evaluation algorithms. Relative weighting of the two strategies is parameterized by w , where 0 reflects purely model-free evaluation and 1, purely model-based. The model-free algorithm implemented is a SARSA(λ) temporal difference algorithm that incrementally updates the value of first-stage stimuli based on both the learned value of a second-stage state and the received reward. The latter is modulated by an eligibility trace parameter λ that only carries value across stages within the same trial. By contrast, the model-based algorithm computes the value of each first-stage choice by multiplying second-stage values by the 70%/30% transition probability. Both algorithms update the second-stage stimulus values the same way, incrementing by the reward-prediction error multiplied by a learning rate α . At each first and second stage decision point, a softmax choice rule is used to assign a probability to each action based on the weighted model-free and model-based values of all available actions; this softmax rule is parameterized by a single inverse temperature parameter (β). A stay bias parameter (p), reflects value-independent perseveration across trials. For each participant's data, the model-based weight

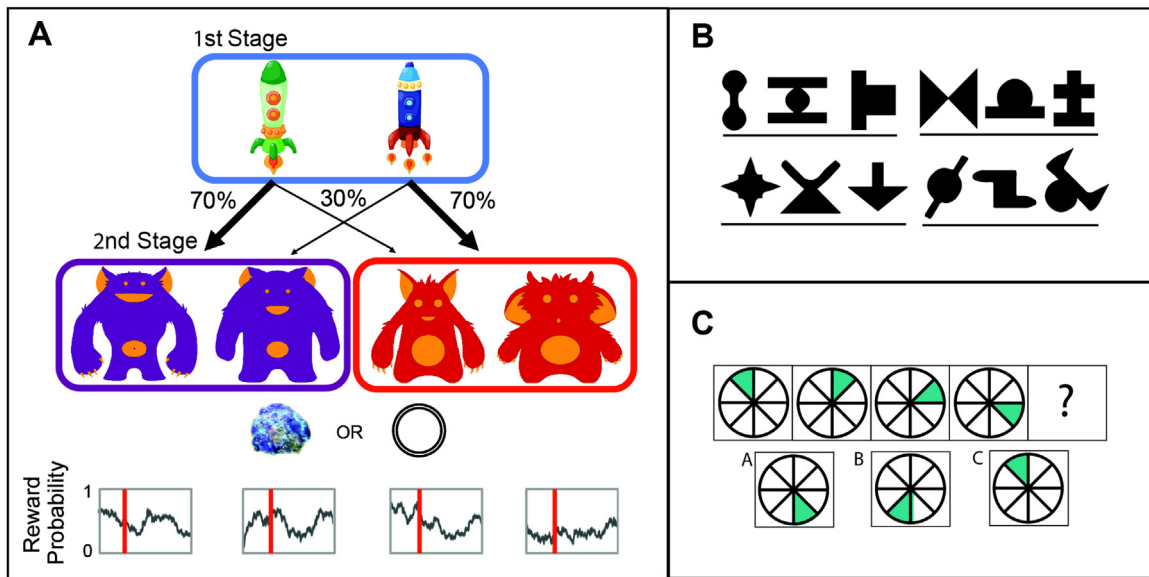


Fig. 1. Task Designs (A) Reinforcement learning task. Each first-stage option (“spaceship”) was associated with one of the second-stage states more frequently (70%) than the other (30%). These transition probabilities were fixed throughout the task. The probability of reward for each second-stage option (“alien”) drifted slowly throughout the 150 trials. (B) Statistical learning task. A continuous stream of stimuli was comprised of four interleaved stimulus triplets. (C) Matrix reasoning task. Example puzzle created to illustrate the type of problems encountered during fluid reasoning task.

(w), learning rate α (α), eligibility parameter λ (λ), softmax inverse temperature parameter (β), and stay bias parameter (p) were estimated simultaneously by maximum a posteriori estimation (Daw et al., 2011).

To evaluate aggregate performance in each categorical age group (i.e., children, adolescents, and adults), we also performed a generalized linear mixed-effects regression analysis on each age group separately using the ‘lme4’ package for the R-statistics language (Bates et al., 2015). First-stage choice (stay or switch from previous trial) was modeled as a function of reward on the previous trial (reward or no reward), transition on the previous trial (rare or common), and the reward-by-transition interaction as fixed effects (Daw et al., 2011; Decker et al., 2016). Model-free and model-based strategies predict different patterns of first-stage choices in the task. Whereas a model-free chooser is likely to repeat rewarded first-stage choices without taking into account the task transition structure, a model-based chooser will be less likely to repeat first-stage choices that are rewarded following a rare transition, and more likely to repeat choices that are unrewarded following a rare transition. Thus, the terms of interest were the fixed effect coefficients of reward (model-free estimate) and the reward-by-transition interaction effect (model-based estimate) for each age group. Additionally, individual adjustments to the fixed intercept (‘random intercept’) and to previous reward, transition, and reward-by-transition interaction terms (‘random slopes’) were determined for each participant. The terms of interest were the fixed-effect coefficients of reward (model-free term) and the reward-by-transition interaction effect (model-based term) for each age group. The first 9 trials for every participant were removed, as were all trials in which the participant did not make both first and second stage choices.

2.3. Statistical learning task

The statistical learning task consisted of two distinct phases, a familiarization phase and a test phase (Schapiro et al., 2014). During the familiarization phase, 12 abstract shapes were presented, one at a time, in a continuous stream without breaks or delays for 4.8 min. Each shape was present for 0.5 s and the shapes were sepa-

rated by a 0.5 s inter-stimulus interval. Participants were instructed to simply watch the stream carefully. Unbeknownst to participants, the sequence of shapes was comprised of 4 distinct triplets (each of the 12 shapes appeared in one triplet only) with a fixed internal order (Fig. 1B). However the triplets were semi-randomly interleaved to form the continuous stream of stimuli such that no triplet was repeated in immediate succession. After the passive viewing stage subjects were then tested on their ability to identify the familiar triplets during a 32-trial test phase. For each trial, subjects were presented with two test triplets, one of which was previously presented during the familiarization phase and the other of which was a foil triplet that was never observed in the presented sequence. Participants were asked to identify which of the two test triplets was more familiar based on the first part of the task. We used the percentage of the familiar triplets that were correctly identified during the test phase as the index of statistical learning ability.

2.4. Fluid reasoning task

Each participant completed two subtests of the Wechsler Abbreviated Scale of Intelligence (WASI), the matrix-reasoning section and the vocabulary section, respectively designed to measure fluid reasoning and crystallized intelligence, (Wechsler, 1999) (Fig. 1C). The subtests were administered according to standard instructions.

The matrix reasoning subsection of the WASI was used as a measure of fluid reasoning. The complete matrix-reasoning section includes 35 puzzles, but children between the ages of 9 and 11 are only presented with the first 32 puzzles. Therefore, to obtain a comparable index of fluid reasoning ability across age groups, we only used participants’ raw scores (number correct) on these first 32 puzzles. While doing so potentially truncated the adolescent and adult scores, it allowed us to evaluate how well all subjects of different ages fared on the same group of puzzles.

To examine whether any observed effects of fluid reasoning were due to a more broadly constructed concept of intelligence, the vocabulary subsection of the WASI was used as a measure of crystallized intelligence. Similarly to the fluid-reasoning index, we used participants’ raw scores (number of points earned) for the first

Table 1

Matrix showing the Pearson correlation coefficients between age and performance on all tasks. Statistically significant relationships denoted in bold. P-values given in parentheses.

	Age	Model-based choice parameter (<i>w</i>)	Statistical learning index	Working memory score	Fluid reasoning score
Age					
Model-based choice parameter (<i>w</i>)	0.30 (0.01)				
Statistical learning index	0.33 (0.007)	0.31 (0.01)			
Working memory score	0.23 (0.12)	0.31 (0.03)	0.09 (0.53)		
Fluid reasoning score	0.53 (<0.0001)	0.41 (<0.001)	0.51 (<0.0001)	0.28 (0.06)	

34 items, which were presented to participants in all age ranges in our study. This scoring method allowed us to evaluate how well all subjects answered the same set of vocabulary questions.

2.5. Working memory task

The listening-recall subtest of the Automated Working Memory Assessment (Alloway, 1999) was used as a measure of working memory, indexing a participant's ability to maintain information despite interference. Participants were read 8 single sentences and 7 pairs of sentences by the researcher. For each single sentence, the participant first listened to the sentence once, then stated whether the sentence was true or false (processing portion) and repeated the last word of the sentence (recall portion). For each pair of sentences, the participant first heard the first sentence, made a true/false declaration, did the same for the second sentence, and then repeated the last words of both sentences in sequential order. Participants were given as much time as needed to complete the processing and recall of each sentence.

For each participant, a recall subscore was tabulated as the number of sequences (single or paired sentences) out of 15 for which the participant correctly recalled the last word(s). A processing subscore was also tabulated as the number of individual sentences for which the participant correctly answered true or false. The participant's working memory score was calculated as the sum of these recall and processing subscores. The processing component, requiring the subject to evaluate the semantic content of the sentence to determine whether it was true or false, serves to interfere with the maintenance of the to-be-recalled information. Poor performance during the processing portion of the task might reflect a failure to engage this competing cognitive process, making recall easier and inflating the recall subscore. By including both the processing subscore and the recall subscore in the measure of working memory, we assessed how well a subject maintained information despite interference.

2.6. Mediation tests

Our mediation analyses were performed using the publicly available causal mediation analysis ('mediate') package for R (Tingley et al., 2014). Non-parametric bootstrap estimates of path coefficients were obtained by random sampling with replacement of 100,000 observations from our sample. Bias-corrected and accelerated (BCa) 95% confidence intervals (DiCiccio and Efron, 1996) were used to calculate two-tailed *p*-values describing significance of the indirect and mediated direct effects.

3. Results

3.1. Learning behavior in two-stage task

We first examined overall performance in the reinforcement-learning task. There was no relationship between age and overall

Table 2

Results of mixed-effects logistic regression quantifying the effects of previous reward and transition type on first-stage choice repetition within each age group. Significant *p*-values (<0.05) denoted in bold.

Predictor	Estimate (SE)	X2 (df = 1)	p-value
Child (N = 19)			
Intercept	0.38 (0.20)	3.54	0.060
Reward	0.26 (0.09)	6.87	0.0088
Transition	−0.07 (0.07)	1.05	0.31
Reward by Transition	0.07 (0.10)	0.45	0.50
Adolescent (N = 22)			
Intercept	0.79 (0.20)	11.61	0.0007
Reward	0.29 (0.11)	6.72	0.0095
Transition	0.02 (0.07)	0.08	0.78
Reward by Transition	0.39 (0.10)	13.14	0.0003
Adult (N = 23)			
Intercept	1.40 (0.18)	28.92	<1e-7
Reward	0.25 (0.09)	6.79	0.0092
Transition	0.06 (0.07)	0.74	0.39
Reward by Transition	0.43 (0.11)	11.41	0.0007

performance on the reinforcement-learning task as indexed by total treasure acquired ($r = 0.18$, $p = 0.14$). Moreover, there was no relationship between total treasure acquired and recruitment of model-based strategy, as indexed by the *w* parameter ($r = -0.09$, $p = 0.48$). Because increased recruitment of model-based learning did not confer an advantage in the ability to earn reward, there was not necessarily an "optimal" strategy in this task. Thus, this task might be better conceived of as indexing an individual's default mode of evaluation, rather than reflecting a cost-benefit calculation of which strategy is best.

We then evaluated whether our cohort showed age-related increases in model-based learning in the two-stage task. We found that the reinforcement-learning *w* parameter significantly and positively correlated with age ($r = 0.30$, $p = 0.01$; Fig. 2, Table 1). Across all subjects, the median and inter-quartile range (IQR) for each reinforcement-learning parameter were: model-based weight (*w*, median = 0.52, IQR = 0.16–0.71), learning rate alpha (α , median = 0.42, IQR = 0.05–0.76), eligibility parameter lambda (λ , median = 0.62, IQR = 0.29–0.90), softmax inverse temperature parameter (β , median = 2.91, IQR = 2.40–4.02), and stay bias parameter (*p*, median = 0.14, IQR = −0.04–0.36).

To quantify the degree of model-based choice at various developmental stages, we conducted a mixed-effects logistic regression analysis within each age group (Table 2). All age groups showed a significant main effect of reward (the model-free signature); however only adolescents ($p < 0.0001$) and adults ($p = 0.0002$), but not children ($p = 0.50$), showed a reward-by-transition interaction effect (the model-based choice signature). These results corroborate a previous finding that evidence of model-based learning is not present in this task during childhood, but emerges in adolescence and increases into early adulthood (Decker et al., 2016).

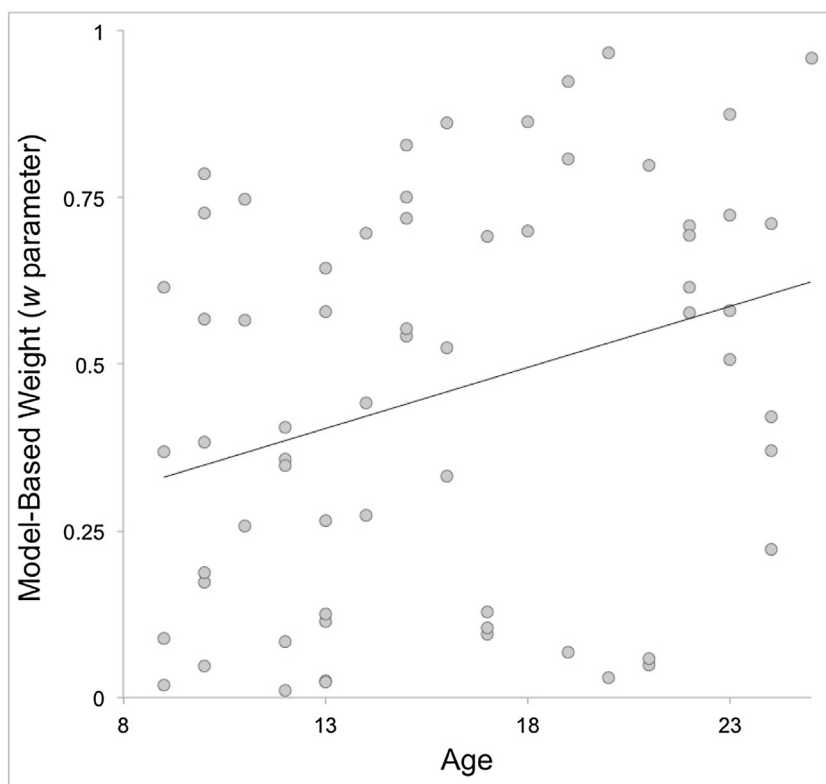


Fig. 2. There was a significant positive correlation between participants' age and the reinforcement-learning w parameter indexing degree of model-based learning ($r = 0.30$, $p = 0.01$).

3.2. Knowledge of transition structure and statistical learning

Participants in all age groups demonstrated similar levels of explicit knowledge of the task structure in their reports after the task ("Which spaceship went mostly to the green planet?") ($X^2 = 3.62$, $df = 2$, $p = 0.16$; 100% (19/19) children, 82% (18/22) adolescents, and 87% (20/23) adults answered correctly).

Examination of response times (RT) following rare versus common transitions can also be interpreted as evidence of a participant's knowledge of the task's transition structure. Slower second-stage responses following rare transitions may reflect a violation of the expectation that the more frequent state transition would occur. Children (paired $t = 2.95$, $df = 18$, $p = 0.0085$), adolescents (paired $t = 2.87$, $df = 21$, $p = 0.0091$), and adults (paired $t = 5.31$, $df = 22$, $p < 0.0001$) all showed significantly slower RTs for rare compared to common trials across the entire task. However, if trials are divided in thirds into early, middle, and late "blocks," only adults (paired $t = 3.93$, $df = 22$, $p = 0.0007$) showed significant RT slowing in the first block of 40 trials (children: paired $t = 1.51$, $df = 18$, $p = 0.15$; adolescents: paired $t = 1.81$, $df = 21$, $p = 0.085$). We then tested whether RT slowing was predictive of greater model-based choice. Such a relationship has previously been shown in adolescents and adults (Decker et al., 2016; Deserno et al., 2015), but did not appear in children (Decker et al., 2016). Corroborating these previous findings, second-stage RT slowing was associated with more model-based choice (indexed by the w parameter from the hybrid RL model) in adults ($r = 0.60$, $p = 0.0025$) and adolescents ($r = 0.56$, $p = 0.0064$), but not children ($r = 0.36$, $p = 0.13$).

3.3. Statistical learning

We found that all age groups showed above chance performance on the statistical learning task (one-sample t -tests: children:

$t = 4.74$, $df = 17$, $p < 0.0001$; adolescents: $t = 8.00$, $df = 21$, $p < 1e-7$; adults: $t = 14.64$, $df = 22$, $p < 1e-12$), but that accuracy did improve with age ($r = 0.33$, $p = 0.0075$). Statistical learning performance also correlated positively with model-based choice, as indexed by the w parameter ($r = 0.31$, $p = 0.01$). A mediation analysis revealed that statistical learning did not mediate the relationship between age and model-based choice (standardized indirect effect 0.08, 95% confidence interval -0.004 – 0.19 , $p = 0.09$). However, it is possible that this trend-level mediation effect would reach significance in a study with a larger sample size.

While participants' response time slowing in the reinforcement-learning task following rare versus common transitions reflected knowledge of the transition structure, this knowledge could have stemmed from experiential learning, or from explicit instruction about the probabilistic transition structure provided to participants in the tutorial. We examined whether participants' RT slowing correlated with statistical learning performance, which might suggest a relationship between RT slowing and the ability to learn underlying statistical regularities through experience. Performance on the statistical learning task correlated positively with second-stage RT slowing in the reinforcement-learning task for rare compared to common trials ($r = 0.41$, $p = 0.0008$).

Collectively, these results suggest that statistical learning may facilitate the construction of a cognitive model of the task in participants of all ages. However, increases in participants' ability to learn experienced environmental statistics did not account for developmental increases in the use of a model-based reinforcement-learning strategy.

3.4. Fluid reasoning

Our measure of fluid reasoning, the raw score on the first 32 questions of the matrix-reasoning subtest of the WASI, increased

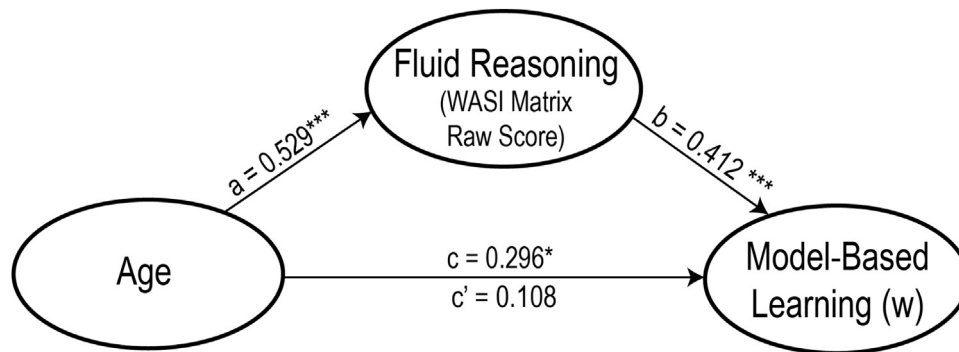


Fig. 3. Fluid reasoning (WASI matrix subscore) fully mediated the relationship between age and model-based strategy. *Denotes $p < 0.05$; ***denotes $p < 0.001$. (single column-fitting image). Path a shows the least-squares regression coefficient of the relationship between age and fluid reasoning. Path b shows the estimated coefficient for the relationship between fluid reasoning and model-based learning. Paths c and c' respectively show coefficients for the effects of age on model-based learning in univariate and multivariate (with fluid reasoning) regressions.

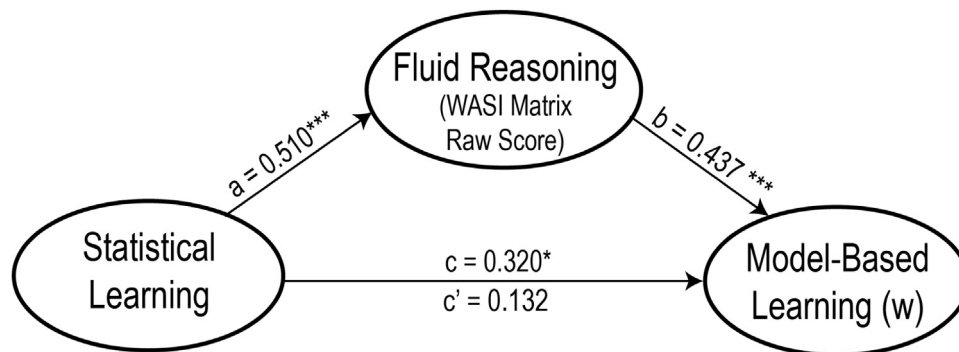


Fig. 4. WASI matrix subscore fully mediated the relationship between statistical learning and model-based learning. *Denotes $p < 0.05$; ***denotes $p < 0.001$. Path a shows the least-squares regression coefficient of the relationship between statistical learning and fluid reasoning. Path b shows the estimated coefficient for the relationship between fluid reasoning and model-based learning. Paths c and c' respectively show coefficients for the effects of statistical learning on model-based learning in univariate and multivariate (with fluid reasoning) regressions.

with age ($r = 0.53$, $p < 0.0001$) and correlated positively with model-based choice, as indexed by the w parameter ($r = 0.41$, $p = 0.001$). A mediation analysis revealed that fluid reasoning fully mediated the relationship between age and model-based choice (standardized indirect effect 0.634, 95% confidence interval 0.198–1.23, $p = 0.01$; standardized direct effect 0.367, 95% confidence interval -0.639 – 1.35 , $p = 0.47$, Fig. 2). These results suggest that age-related increases in fluid reasoning are a critical cognitive factor underlying the previously observed association between age and increased recruitment of model-based learning (Decker et al., 2016).

To determine whether the observed effect was specific to fluid reasoning or could be attributed to intelligence more broadly, we next tested whether vocabulary also mediated the relationship between age and model-based choice. Vocabulary scores were positively correlated with both age ($r = 0.42$, $p < 0.001$) and model-based choice, as indexed by the reinforcement learning w parameter ($r = 0.28$, $p = 0.032$). However, mediation analysis showed that vocabulary scores did not mediate the relationship between age and model-based choice (standardized indirect effect 0.077, 95% confidence interval -0.034 to 0.172 , $p = 0.15$), suggesting that the effect of fluid reasoning was specific, rather than relating to crystallized intelligence or IQ more generally.

As we hypothesized that statistical learning may promote the construction of mental models of transition structure, and that fluid reasoning might facilitate the recruitment of such models, we next tested whether fluid reasoning mediated the relationship between statistical learning and model-based choice. Fluid reasoning fully mediated the relationship between statistical learning and model-based choice (standardized indirect effect 0.614, 95% confidence

interval 0.213–1.197, $p = 0.01$; standardized direct effect 0.430, 95% confidence interval -0.466 to 1.400 , $p = 0.36$, Fig. 3). Furthermore, this mediation was directionally specific: statistical learning did not fully or partially mediate the relationship between fluid reasoning and model-based choice (standardized indirect effect 0.292, 95% confidence interval -0.292 to 1.042 , $p = 0.36$). Collectively, these results suggest that, independent of age, fluid reasoning accounts for the relationship between an individual's statistical learning ability and their recruitment of model-based learning.

In addition, fluid reasoning fully mediated the relationship between age and statistical learning (standardized indirect effect 0.227, 95% confidence interval 0.107–0.435, $p < 0.001$; standardized direct effect 0.14, 95% confidence interval -0.150 to 0.385 , $p = 0.31$, Fig. 4). This mediation was directionally specific: statistical learning did not mediate the relationship between fluid reasoning and age (standardized indirect effect 0.071, 95% confidence interval -0.075 to 0.213 , $p = 0.31$). These analyses suggest that fluid reasoning contributes to age-related increases in statistical learning ability.

3.5. Working memory

Participants performed well on our measure of working memory, the listening recall subtest of the Automated Working Memory Assessment, with 45% of participants, spanning all age groups, exhibiting ceiling-level performance. This ceiling effect limited our ability to use this measure in order to clarify the relationship between working memory and age-related increases in the recruitment of model-based learning. Nonetheless, we observed that working memory correlated positively with model-based choice

($r=0.31$, $p=0.03$), but not with age ($r=0.23$, $p=0.12$, Table 1). We also observed a trending positive relationship between our working memory and fluid reasoning measures ($r=0.28$, $p=0.06$, Table 1), consistent with the extensive previous examinations of this relationship (Colom et al., 2008; Kane and Engle, 2002). However, these relationships should be interpreted with caution given our failure to obtain a robust index of individual differences in working memory performance.

We also examined whether the AWMA recall subscore alone might provide a more robust measure of working memory. To test whether variability in the recall subscore might be greater than for the composite measure, we calculated the coefficient of variation (the mean-scaled standard deviation) for both the recall subscore (0.08) and composite score (0.07) and found that the variability was comparable for both measures. However, the percentage of participants who showed ceiling performance increased to 69% using the recall subscore alone. This suggests that the recall subscore also failed to provide a valid metric of working memory performance in the sample. Furthermore, we did not find a relationship between the recall subscore and age ($r=0.20$, $p=0.17$) or between the recall subscore and the reinforcement-learning w parameter ($r=0.15$, $p=0.31$), indicating no possibility of statistical mediation.

4. Discussion

In this study, we sought to elucidate the cognitive components that underlie the developmental emergence of model-based learning. Replicating previous findings in this sequential reinforcement-learning task (Decker et al., 2016), we found that whereas model-free learning was evident across our developmental sample, model-based choice exhibited a protracted maturational trajectory, only becoming evident in adolescence and continuing to strengthen into adulthood. We examined whether developmental changes in statistical learning ability, working memory, and fluid reasoning might contribute to the increased recruitment of model-based choice with age. We found that statistical learning performance was evident in children and improved with age. However, these improvements in statistical learning did not account for age-related increases in model-based choice. In contrast, fluid reasoning increased with age and significantly mediated the relationship between age and model-based learning. Collectively, these results suggest that the ability to integrate distinct learned associations, and not merely the acquisition of those associations, is a critical cognitive component underlying the gradual development of model-based choice.

Although children did not show evidence of model-based learning, they demonstrated knowledge of the task transition structure. Children, like adolescents and adults, could explicitly describe the task structure and were also slower to respond following rare transitions, reflecting sensitivity to these less frequent outcomes. Notably, whereas adults' response time sensitivity was apparent early in the task, this sensitivity only emerged later in children. Adults are able to rapidly incorporate explicit instruction to inform their actions (Cole et al., 2013), and may have used the task description provided in our tutorial to scaffold a cognitive model of the probabilistic transition structure. In contrast, younger participants, who tend to rely on experiential over instructed knowledge (Decker et al., 2015), may have had greater difficulty recruiting this instruction to inform choices. Thus, providing instruction may have facilitated adults' recruitment of a model-based strategy. Younger participants may instead have learned the task structure predominantly through the experience they accumulated over many trials. This proposal is consistent with the observed correlation between response time slowing and participants' performance on the statistical learning task, in which sequential regularities were learned

solely through experience. Future studies might test whether children exhibit model-based choice if this learned task structure knowledge grows more "crystallized" through extensive practice.

Performance on the Wechsler Abbreviated Scale of Intelligence matrix-reasoning subtest, an index of fluid-reasoning ability, significantly mediated the relationship between age and model-based choice. A major component of fluid reasoning is the ability to identify associations between mental representations across distinct "dimensions", often referred to as relational integration (Wright et al., 2008). As the number of relevant dimensions increases, the more complex the relation becomes (Christoff et al., 2001). For example, low-level integration may involve representing a simple characteristic of an object (e.g., *shape A is a circle*), while higher levels of integration may involve identifying relations between properties of multiple objects (e.g., *shape B is also a circle, but a different color*) or assessing the relationship between multiple relations (e.g., *shape A is to B as shape C is to D*). Like fluid-reasoning tasks, model-based choice requires the integration of learned relationships across multiple dimensions. In our sequential reinforcement-learning task, a model-based chooser must be able to prospectively integrate the transition probabilities between the first- and second-stage states with the reward probabilities associated with each second-stage stimuli. In fluid-reasoning puzzles that require considering two or more joint relations, children have been found to select answers at the same speed as adults, but with less accuracy (Crone et al., 2009; Vodegel Matzen et al., 1994). These findings suggest that children may not consider all the relevant dimensions of the problem before selecting an answer. Children in our reinforcement-learning task may similarly not recognize that recruiting transition knowledge at the first stage influences their later options, and therefore fail to integrate this knowledge into their evaluation at the first-stage.

Working memory plays a significant role in both fluid reasoning (Kane and Engle, 2002) and model-based choice (Otto et al., 2013a, 2013b). A large body of research has shown that working memory improves with age (De Luca et al., 2003; Tamnes et al., 2013). In this study, nearly half of the participants, across all age groups, in whom we assessed working memory performed at ceiling. Thus, we failed to obtain a reliable index of working memory, precluding the ability to clearly characterize its contribution to developmental changes in reinforcement-learning strategy. Nonetheless, we observed a significant relationship between working memory and model-based choice, suggesting that age-related improvements in working memory may contribute to the development of model-based learning. Fluid reasoning, which mediated improvements in model-based choice in our sample, also depends on working memory ability. Thus, future studies, employing more robust assessments of working memory, will be necessary to dissociate the contributions of working memory and fluid reasoning to the recruitment of model-based learning.

The neurocircuitry underlying the development of model-based learning has not been directly characterized. However, the neural substrates underlying the cognitive processes examined in this study have been previously explored. Statistical learning depends on medial temporal lobe structures, including the hippocampus (Davachi and DuBrow, 2015; Preston et al., 2004; Schapiro and Turk-Browne, 2015). Recent findings suggest that developmental improvements in statistical learning parallel hippocampal structural development (Schlichting et al., 2016). To inform goal-directed actions, these learned sequential representations must be integrated with learned associations between stimuli, actions, and past rewards, which depend in part on contributions from the striatum (Balleine and O'Doherty, 2010). Working memory and relational integration have been shown to depend on dorsolateral (Curtis and D'Esposito, 2003) and rostralateral (Wright et al., 2008) prefrontal cortical regions respectively. Cortical maturation typ-

ically proceeds from posterior to anterior cortical regions, with the dorsolateral prefrontal cortex being one of the latest maturing regions (Gogtay et al., 2004; Shaw et al., 2008). This developmental trajectory mirrors the proposed organizational hierarchy of the prefrontal cortex, in which increasingly complex and abstract representations recruit more anterior regions (Badre and D'Esposito, 2007; Dixon and Christoff, 2014). Thus, younger individuals, for whom rostral and dorsolateral prefrontal regions are not yet mature, may have greater difficulty integrating and recruiting the multi-dimensional cognitive representations on which model-based learning depends. More broadly, this literature suggests that elements of the integrated prefrontal-hippocampal-striatal circuits underpinning key component processes of model-based learning exhibit distinct, and often protracted, maturational trajectories. Direct examination of the age-related changes in these circuits will be necessary to further elucidate their specific functional roles in the developmental emergence of model-based choice.

In this study, we found that the fluid-reasoning abilities within a cohort of children, adolescents, and adults accounted for age-related changes in their recruitment of model-based learning. Fluid reasoning is a construct originating from the field of psychometric intelligence (Cattell, 1987), which typically employs tasks that are substantially different from those employed within cognitive psychology and neuroscience to assess related constructs such as working memory and other executive functions. While there is debate regarding the degree to which these overlapping constructs are indeed dissociable, (Friedman et al., 2006; Decker et al., 2007), our findings are consistent with others in the literature observing correlations between these cognitive constructs and reward-guided behaviors (Shamosh et al., 2008). We propose that such correlations reflect a mechanistic relationship, in which these cognitive processes provide a foundation for the evaluative computations that inform motivated behavior. This interpretation, supported by our present findings, suggests a direct link between the protracted development of these cognitive abilities (Ferrer et al., 2009) and the marked developmental changes in reward-related decision-making (Hartley and Somerville, 2015).

Funding

This work was supported by the National Institute on Drug Abuse (R03-DA038701), the Alice Bohmfalk Charitable Trust, an Esther Katz Rosen Fund grant from the American Psychological Foundation, and a generous gift from the family of Mortimer D. Sackler.

Conflict of interest

None.

Acknowledgements

We thank Lindsay Hunter for assistance with data collection, Johannes Decker for assistance with data analysis, and Nicholas Turk-Browne for sharing the statistical learning task.

References

- Alloway, T.P., 1999. *Automated Working: Memory Assessment: Manual*. Pearson.
- Amso, D., Davidow, J., 2012. The development of implicit learning from infancy to adulthood: item frequencies, relations, and cognitive flexibility. *Dev. Psychobiol.* 54, 664–673, <http://dx.doi.org/10.1002/dev.20587>.
- Badre, D., D'Esposito, M., 2007. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J. Cogn. Neurosci.* 19 (12), 2082–2099.
- Balleine, B.W., O'Doherty, J.P., 2010. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35, 48–69, <http://dx.doi.org/10.1038/npp.2009.131>.
- Bates, D., Maechler, M., Bolker, B., Walker, S., (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–9, URL: <https://CRAN.R-project.org/package=lme4>.
- Cattell, R.B., 1987. *Intelligence: Its Structure, Growth and Action: Its Structure, Growth and Action*, vol. 35. Elsevier.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J.K., Holyoak, K.J., Gabrieli, J.D., 2001. Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage* 14, 1136–1149, <http://dx.doi.org/10.1006/nimg.2001.0922>.
- Cole, M.W., Laurent, P., Stocco, A., 2013. Rapid instructed task learning: a new window into the human brain's unique capacity for flexible cognitive control. *Cogn. Affect. Behav. Neurosci.* 13, 1–22, <http://dx.doi.org/10.3758/s13415-012-0125-7>.
- Colom, R., Abad, F.J., Quiroga Á, M., Shih, P.C., Flores-Mendoza, C., 2008. Working memory and intelligence are highly related constructs, but why? *Intelligence* 36 (6), 584–606.
- Crone, E.A., Wendelken, C., Van Leijenhorst, L., Honomichl, R.D., Christoff, K., Bunge, S.A., 2009. Neurocognitive development of relational reasoning. *Dev. Sci.* 12, 55–66, <http://dx.doi.org/10.1111/j.1467-7687.2008.00743.x>.
- Curtis, C.E., D'Esposito, M., 2003. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423, [http://dx.doi.org/10.1016/S1364-6613\(03\)00197-9](http://dx.doi.org/10.1016/S1364-6613(03)00197-9).
- D'Esposito, M., Postle, B.R., 2015. The Cognitive Neuroscience of Working Memory 66, 115–142. 10.1146/annurev-psych-010814-015031. THE.
- Davachi, L., DuBrow, S., 2015. How the hippocampus preserves order: the role of prediction and context. *Trends Cogn. Sci.* 19, 92–99.
- Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711.
- 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, 1204–1215, <http://dx.doi.org/10.1016/j.neuron.2011.02.027>.
- De Luca, C.R., Wood, S.J., Anderson, V., Buchanan, J.A., Proffitt, T.M., Mahony, K., Pantelis, C., 2003. Normative data from the CANTAB. I: development of executive function over the lifespan. *J. Clin. Exp. Neuropsychol.* 25 (2), 242–254.
- Decker, S.L., Hill, S.K., Dean, R.S., 2007. Evidence of construct similarity in executive functions and fluid reasoning abilities. *Int. J. Neurosci.* 117 (6), 735–748.
- Decker, J.H., Lourenco, F.S., Doll, B.B., Hartley, C.A., 2015. Experiential reward learning outweighs instruction prior to adulthood. *Cogn. Affect. Behav. Neurosci.* 15, 310–320, <http://dx.doi.org/10.3758/s13415-014-0332-5>.
- Decker, J.H., Otto, A.R., Daw, N.D., Hartley, C.A., 2016. From creatures of habit to goal-directed learners: tracking the developmental emergence of model-based reinforcement learning. *Psychol. Sci.*, <http://dx.doi.org/10.1177/0956797616639301>.
- Deserno, L., Huys, Q.J.M., Boehme, R., Buchert, R., Heinze, H., 2015. Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proc. Natl. Acad. Sci. U. S. A.* 112, 1595–1600, <http://dx.doi.org/10.1073/pnas.1417219112>.
- DiCiccio, T.J., Efron, B., 1996. Bootstrap confidence intervals. *Stat. Sci.* 11, 189–228, <http://dx.doi.org/10.1214/ss/1032280214>.
- Dixon, M.L., Christoff, K., 2014. The lateral prefrontal cortex and complex value-based learning and decision making. *Neurosci. Biobehav. Rev.* 45, 9–18, <http://dx.doi.org/10.1016/j.neubiorev.2014.04.011>.
- Ferrer, E., O'Hare, E.D., Bunge, S.A., 2009. Fluid reasoning and the developing brain. *Front. Neurosci.* 3, 46–51, <http://dx.doi.org/10.3389/neuro.01.003.2009>.
- Fiser, J., Aslin, R.N., 2002. Statistical learning of new visual feature combinations by infants. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15822–15826, <http://dx.doi.org/10.1073/pnas.232472899>.
- Friedman, N.P., Miyake, A., Corley, R.P., Young, S.E., DeFries, J.C., Hewitt, J.K., 2006. Not all executive functions are related to intelligence. *Psychol. Sci.* 17 (2), 172–179.
- Fry, A.F., Hale, S., 1996. Processing speed, working memory, and fluid intelligence: evidence for a developmental cascade. *Psychol. Sci.* 7, 237–241.
- Glascher, J., Daw, N., Dayan, P., O'Doherty, J.P., 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 585–595, <http://dx.doi.org/10.1016/j.neuron.2010.04.016>.
- Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent 3rd, T.F., Herman, D.H., Clasen, L.S., Toga, A.W., Rapoport, J.L., Thompson, P.M., 2004. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8174–8179, <http://dx.doi.org/10.1073/pnas.0402680101>.
- Hartley, C.A., Somerville, L.H., 2015. The neuroscience of adolescent decision-making. *Curr. Opin. Behav. Sci.* 5, 108–115.
- Kane, M.J., Engle, R.W., 2002. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon. Bull. Rev.* 9, 637–671, <http://dx.doi.org/10.3758/BF03196323>.
- Klossek, U.M.H., Russell, J., Dickinson, a., 2008. The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. *J. Exp. Psychol. Gen.* 137, 39–51, <http://dx.doi.org/10.1037/0096-3445.137.1.39>.

- McArdle, J.J., Ferrer-Caja, E., Hamagami, F., Woodcock, R.W., 2002. Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Dev. Psychol.* 38, 115–142, <http://dx.doi.org/10.1037/0012-1649.38.1.115>.
- Otto, A.R., Gershman, S.J., Markman, A.B., Daw, N.D., 2013a. The curse of planning dissecting multiple reinforcement-Learning systems by taxing the central executive. *Psychol. Sci.* 24, 751–761, <http://dx.doi.org/10.1177/0956797612463080>.
- Otto, A.R., Raio, C.M., Chiang, A., Phelps, E.A., Daw, N.D., 2013b. Working-memory capacity protects model-based learning from stress. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20941–20946, <http://dx.doi.org/10.1073/pnas.1312011110>.
- Piaget, J., 1954. *The Construction of Reality in the Child*. Basic Books.
- Preston, A.R., Shrager, Y., Dudukovic, N.M., Gabrieli, J.D.E., 2004. Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus* 14, 148–152, <http://dx.doi.org/10.1002/hipo.20009>.
- Schapiro, A., Turk-Browne, N.B., 2015. Statistical learning. *Brain Mapp.* 3, 501–506, <http://dx.doi.org/10.1016/B978-0-12-397025-1.00276-1>.
- Schapiro, A.C., Gregory, E., Landau, B., McCloskey, M., Turk-Browne, N.B., 2014. The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci.* 26, 1736–1747, <http://dx.doi.org/10.1162/jocn.a.00578>.
- Schlichting, M.L., Guarino, K.F., Schapiro, A.C., Turk-Browne, N.B., Preston, A.R., 2016. Hippocampal structure predicts statistical learning and associative inference abilities during development. *J. Cogn. Neurosci.*
- Shamosh, N.A., DeYoung, C.G., Green, A.E., Reis, D.L., Johnson, M.R., Conway, A.R., Gray, J.R., 2008. Individual differences in delay discounting relation to intelligence, working memory, and anterior prefrontal cortex. *Psychol. Sci.* 19 (9), 904–911.
- Shaw, P., Kabani, N.J., Lerch, J.P., Eckstrand, K., Lenroot, R., Gogtay, N., Greenstein, D., Clasen, L., Evans, A., Rapoport, J.L., Giedd, J.N., Wise, S.P., 2008. Neurodevelopmental trajectories of the human cerebral cortex. *J. Neurosci.* 28, 3586–3594, <http://dx.doi.org/10.1523/JNEUROSCI.5309-07.2008>.
- Tamnes, C., Walhovd, K., Grydeland, H., 2013. Longitudinal working memory development is related to structural maturation of frontal and parietal cortices. *J. Cogn. Neurosci.* 25, 1611–1623, <http://dx.doi.org/10.1162/jocn.a.00434>.
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., Imai, K., 2014. Mediation: r package for causal mediation analysis. *J. Stat. Softw.* 59 (5), 1–38 (URL <http://www.jstatsoft.org/v59/i05/>).
- Turk-Browne, N.B., Jungé, J., Scholl, B.J., 2005. The automaticity of visual statistical learning. *J. Exp. Psychol. Gen.* 134, 552–564, <http://dx.doi.org/10.1037/0096-3445.134.4.552>.
- Vodegel Matzen, L.B.L., van der Molen, M.W., Dudink, A.C.M., 1994. Error analysis of raven test performance. *Pers. Individ. Dif.* 16, 433–445, [http://dx.doi.org/10.1016/0191-8869\(94\)90070-1](http://dx.doi.org/10.1016/0191-8869(94)90070-1).
- Wechsler, D., 1999. *Wechsler Abbreviated Scale of Intelligence (WASI)*. Psychological Corporation, London.
- Wright, S.B., Matlen, B.J., Baym, C.L., Ferrer, E., Bunge, S.A., 2008. Neural correlates of fluid reasoning in children and adults. *Front. Hum. Neurosci.* 1, 8, <http://dx.doi.org/10.3389/neuro.09.008.2007>.
- Zelazo, P.D., Frye, D., Rapus, T., 1996. An age-related dissociation between knowing rules and using them. *Cogn. Dev.* 11, 37–63, [http://dx.doi.org/10.1016/S0885-2014\(96\)90027-1](http://dx.doi.org/10.1016/S0885-2014(96)90027-1).