Research Article

Psychological Science
1–11

DOI: 10.1177/0956797616639301
pss.sagepub.com

From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning

Johannes H. Decker[1], A. Ross Otto[2], Nathaniel D. Daw[3,4], and Catherine A. Hartley[1]

[1]Sackler Institute for Developmental Psychobiology, Weill Cornell Medical College; [2]Center for Neural Science, New York University; [3]Princeton Neuroscience Institute, Princeton University; and [4]Department of Psychology, Princeton University

Abstract

Theoretical models distinguish two decision-making strategies that have been formalized in reinforcement-learning theory. A model-based strategy leverages a cognitive model of potential actions and their consequences to make goal-directed choices, whereas a model-free strategy evaluates actions based solely on their reward history. Research in adults has begun to elucidate the psychological mechanisms and neural substrates underlying these learning processes and factors that influence their relative recruitment. However, the developmental trajectory of these evaluative strategies has not been well characterized. In this study, children, adolescents, and adults performed a sequential reinforcement-learning task that enabled estimation of model-based and model-free contributions to choice. Whereas a model-free strategy was apparent in choice behavior across all age groups, a model-based strategy was absent in children, became evident in adolescents, and strengthened in adults. These results suggest that recruitment of model-based valuation systems represents a critical cognitive component underlying the gradual maturation of goal-directed behavior.

Keywords

cognitive development, reinforcement learning, decision making, open data

Received 7/30/15; Revision accepted 2/24/16

Learning to select actions that yield the best outcomes is a lifelong challenge. From even very young ages, children demonstrate competence in making many simple value-based decisions. However, some aspects of decision making also exhibit qualitative changes across development. Younger individuals often persist with actions that no longer yield beneficial outcomes, and such perseveration decreases with age (Klossek, Russell, & Dickinson, 2008; Piaget, 1954). Children and adolescents often make seemingly shortsighted choices that prioritize immediate gains over longer-term rewards (Mischel, Shoda, & Rodriguez, 1989). Such choices have been proposed to reflect regulatory failures in which insufficient executive control leads to prepotent action or prioritization of hedonically alluring outcomes over more valuable alternatives (Posner & Rothbart, 2000). Indeed, executive functions such as cognitive control and working memory improve markedly from childhood into adulthood (Diamond, 2006). Although typically studied in controlled isolation, these executive processes interact to inform people's choices in more ecologically relevant behavioral contexts. Recent findings in adults suggest that integrated executive functioning provides a cognitive foundation for complex decision computations that alter the manner in which reward-related actions are evaluated (Otto, Gershman, Markman, & Daw, 2013; Otto,

Corresponding Author:
Catherine A. Hartley, Weill Cornell Medical College, Sackler Institute for Developmental Psychobiology, 1300 York Ave., Box 140, New York, NY 10065
E-mail: cah2031@med.cornell.edu

Skatova, Madlon-Kay, & Daw, 2015). This work suggests that normal cognitive development may simultaneously give rise to changes in the evaluative process through which an individual determines which actions are best.

Theoretical models distinguish two types of evaluative processes that can inform one's choices (Daw, Niv, & Dayan, 2005). A slower, deliberative, goal-directed process compares potential actions and their likely consequences to identify the action most likely to obtain a desired outcome. In contrast, a more rapid and automatic habitual process links rewarded actions to associated cues and contexts, enabling reflexive repetition of previously successful behaviors. A large psychological and neuroscientific literature provides support for these distinct evaluative strategies (Balleine & O'Doherty, 2009; Dickinson, 1985; Doll, Simon, & Daw, 2012). Two classes of reinforcement-learning algorithms are proposed to approximate their underlying neural computations and to capture their key behavioral properties (Daw et al., 2005). *Model-based* algorithms select actions via a flexible but computationally demanding process of searching a cognitive model of potential state transitions and outcomes. In contrast, *model-free* algorithms recruit trial-and-error feedback to efficiently update a cached action value associated with a stimulus. Adaptive control of behavior involves a fluid and contextually sensitive balance between these dissociable learning systems. Whereas model-free learning promotes the execution of well-honed behavioral routines without forethought or attention, model-based learning enables flexible adaptation of behavior to the dynamic state of the world.

In adulthood, model-free and model-based systems are proposed to operate in parallel, competing for control over behavior (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Dickinson, 1985). Reliance on a given strategy appears to be sensitive to the cognitive and affective demands placed on the individual (Otto, Gershman, et al., 2013; Otto, Raio, Chiang, Phelps, & Daw, 2013). From childhood into adulthood, the prefrontal-subcortical neurocircuitry implicated in model-based learning (Daw et al., 2005) undergoes substantial structural and functional changes (Somerville & Casey, 2010), suggesting that the relative reliance on these two forms of learning might change markedly with age. However, the developmental trajectory of these action-selection strategies has not yet been examined. In the current study, we examined the extent to which children, adolescents, and adults exhibited the behavioral signatures of model-free and model-based strategies, using a two-stage reinforcement-learning task designed to distinguish these two forms of learning. Whereas a model-free strategy was evident across all age groups, a model-based influence on choice emerged only in adolescents and continued to increase in adults. Collectively, these results suggest that the recruitment of model-based evaluative processes emerges gradually with age, highlighting a critical cognitive component underlying the development of goal-directed decision making.

Method

Participants

Thirty children (age range = 8–12 years), 28 adolescents (age range = 13–17 years), and 22 adults (age range = 18–25 years) completed the task. On the basis of a previous study (Eppinger, Walter, Heekeren, & Li, 2013) that found a large effect (η^2 = .2) of aging on model-based evaluation in adulthood, we expected that our target sample size of 60 (20 participants per age group) would achieve approximately 90% power to detect a true effect of a comparable size, assuming an α of .05. We recruited higher numbers of children and adolescents because we expected higher rates of attrition in these age groups. The final sample included 59 participants: 20 children (11 females; mean age = 9.80 years, *SD* = 1.54 years), 20 adolescents (12 females; mean age = 15.35 years, *SD* = 1.39 years), and 19 adults (11 females; mean age = 21.63 years, *SD* = 2.03). For details on the exclusion criteria, see the next section. All participants provided written informed consent according to the procedures of the Weill Cornell Medical College institutional review board. All participants were compensated $30 regardless of their performance.

Reinforcement-learning task (spaceship task)

We adapted a sequential learning task from Daw et al. (2011) that was designed to dissociate model-free and model-based learning strategies, to use a child-friendly narrative, and to be engaging for a developmental cohort. Before the task, all participants completed a tutorial that conveyed the task cover story and introduced key concepts, such as probabilistic rewards and transitions, via a series of interactive example trials. The tutorial was automated to ensure that all participants received equivalent information, and it concluded with an instruction summary using simple child-friendly terminology. All participants indicated verbally that they understood the instructions before starting the task.

Participants were tasked with collecting "space treasure" (Fig. 1a). First, they chose between two spaceship stimuli (first-stage choice). Each spaceship traveled more frequently to one planet than to the other (70% versus 30%). For example, the blue spaceship had a 70% probability of leading to the red planet (the common transition) and a 30% probability of leading to the purple planet (the rare transition). The green spaceship had the

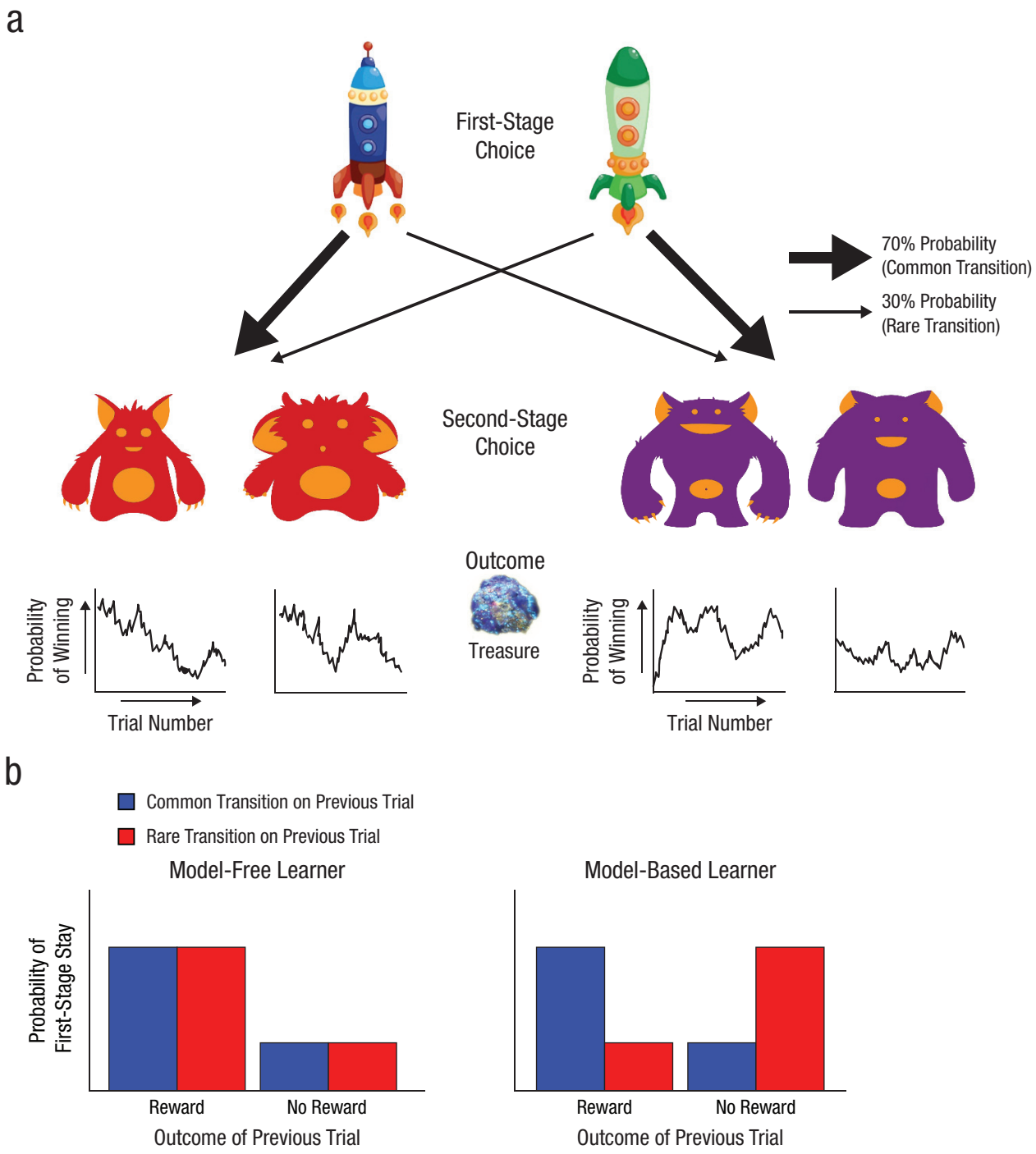

Fig. 1. Design of the sequential spaceship task (a) and idealized model-free and model-based behavior (b). On each trial, participants chose between two spaceships (first-stage choice), which was followed by a probabilistic transition to a red planet or a purple planet. Then participants chose between two aliens (second-stage choice) and were rewarded with space treasure or not. The probability of winning space treasure is presented as a function of trial for each alien. The bar graphs show, for idealized model-free and model-based learners, the probability of making the same choice on the next trial (i.e., a first-stage stay) as a function of the outcome and transition type (common or rare) of the previous trial.

opposite probabilities (i.e., 70% chance of the purple planet and 30% chance of the red planet). On each planet, participants chose between two alien-creature stimuli (second-stage choice). They were then rewarded with a picture of space treasure or with nothing (an empty circle) according to a slowly drifting probability

(between 0.2 and 0.8). These shifting reward probabilities encouraged participants to explore different choices throughout the task to maximize rewards. Participants had 3 s to make each choice, followed by a 1-s animation, 1 s of reward feedback, and a 1-s intertrial interval. The full game consisted of 200 trials in four blocks separated by breaks.

Data from 1 child and 1 adolescent were excluded because of inattentiveness (e.g., looking away from the screen, closing their eyes), and 1 adult was excluded for making the same choice on every trial throughout the task. In addition, we used two criteria to exclude participants whose behavior was inconsistent with an intention to obtain rewards in the task: (a) The proportion of first-stage stay decisions after common transitions had to be at least .1 greater after rewarded trials than after unrewarded trials (1 child, 4 adolescents, and 1 adult were excluded), and (b) participants who encountered a second-stage state that was rewarded on the previous trial were required to repeat the rewarded choice at least 55% of the time (8 children, 3 adolescents, and 1 adult were excluded). One adolescent failed both criteria but is counted here only among the exclusions for the first criterion. Under these criteria, participants were required to appear to be pursuing reward but whether they did so via a model-free or model-based strategy was irrelevant.

Critically, this task structure enables dissociation of the relative recruitment of model-free and model-based learning strategies. Whereas a model-based chooser would use a cognitive model of the transition types and outcomes in the task to select actions, a model-free chooser simply repeats previously rewarded actions (Fig. 1b). Thus, how a previous trial influences the subsequent first-stage choice depends on one's learning strategy. For example, consider a trial in which one chooses the blue spaceship, makes a rare transition to the purple planet, chooses an alien, and is rewarded. A model-free learner would be likely to repeat the previous first-stage choice (i.e., the blue spaceship), regardless of the transition type that led to the reward (i.e., there is a main effect of reward). In contrast, a model-based chooser—taking into account the state-transition structure—would be likely to switch to the green spaceship, increasing the likelihood of returning to that rewarded state (i.e., there is a reward-by-transition-type interaction effect).

Behavioral analysis

Logistic regression analysis of this task has been described previously (Daw et al., 2011; Otto, Gershman, et al., 2013). In brief, a generalized linear mixed-effects regression analysis of group behavior data was performed using the lme4 package (Version 1.1-8; Bates, Maechler, Bolker, & Walker, 2015) for the R software environment (Version 3.3.3; R Development Core Team, 2015). First-stage choice (stay or switch from previous trial) was modeled by independent predictors of previous reward (reward or no reward), previous transition type (rare or common), age (z-score transformed), and all two-way and three-way interactions as fixed effects, as well as per-participant random adjustment to the fixed intercept (random intercept) and per-participant adjustment to previous reward, transition-type, and reward-by-transition-type interaction terms (random slopes). The terms of interest were the main effect of reward (i.e., the model-free term), a reward-by-transition-type interaction effect (i.e., the model-based term), and the reward-by-age and reward-by-transition-type-by-age interaction effects. The first 9 trials for every subject were removed, as were trials in which an individual failed to make a first- or second-stage choice (median number of trials removed: for children, 3.5; for adolescents, 0.5; for adults, 0). In addition, this analysis was performed separately for each age group by removing the age and age-interaction terms. Response time data for the second-stage actions were analyzed similarly using a linear mixed-effects analysis with current transition type and age as independent predictors. Finally, the relationship between individual random-effects estimates of the model-based term and response time difference was examined through bivariate correlation.

We also fit subjects' choices using a reinforcement model similar to the hybrid model described in Otto, Gershman, et al. (2013). This model allows participant's choices to take into account the entire preceding history of rewards and assumes that choices are determined by a weighted combination of model-free and model-based values. We used hierarchical Bayesian model-fitting techniques to derive individual model-based and model-free weight parameters. For each parameter, we computed a 95% confidence interval (CI). If the entire CI fell above or below zero, we concluded that the parameter of interest was significantly positive or negative, respectively, with 95% confidence. Details of the model-fitting procedure are provided in the Supplemental Material available online.

Results

Learning behavior

We assessed participants' recruitment of these two learning strategies by examining the effects of transition type (common or rare) and reward on their subsequent first-stage choices (stay or switch). The qualitative pattern of these choices within the child, adolescent, and adult categorical age groups is shown in Figure 2a. Whereas children's choices closely resembled the pattern of an idealized model-free chooser, adolescents and adults exhibited a mixture of choice strategies, as has been observed previously in adults (Daw et al., 2011; Otto, Gershman, et al., 2013). We

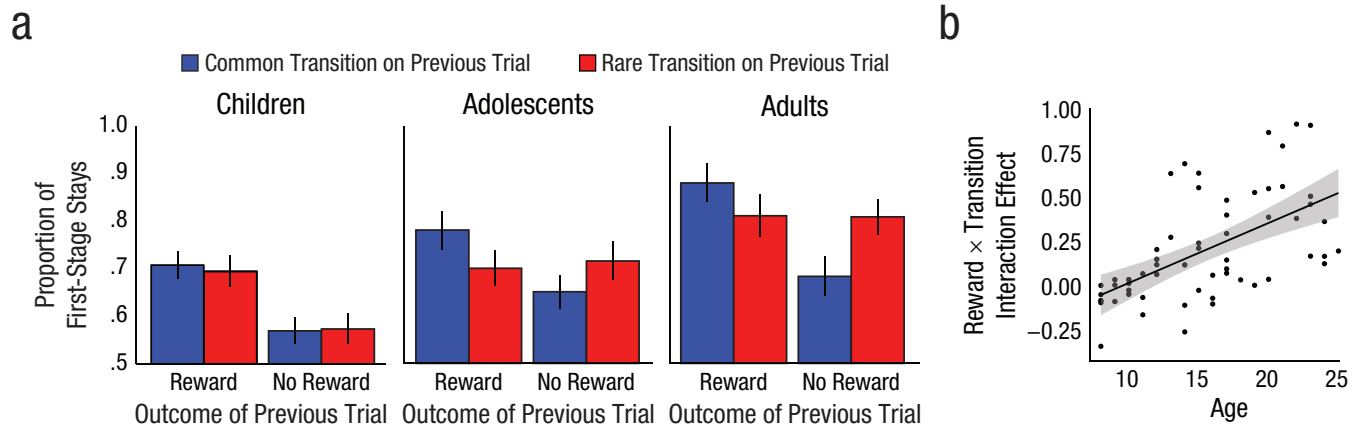

Fig. 2. First-stage choice behavior by age. The proportion of first-stage stay choices is graphed as a function of outcome of the previous trial for each age group (a), separately for trials following common and rare transitions. The error bars represent ±1 *SEM*. The scatterplot (with best-fitting regression line) shows the relationship between the reward-by-transition-type interaction effect (the model-based effect estimates) and age (b). The model-based effect is plotted as the fixed plus the random effects from a regression model with age excluded. The gray area represents ±1 *SEM*.

conducted a mixed-effects logistic regression analysis within each categorical age group to quantify these age-related choice patterns (Table 1). Children ($p = .0006$), adolescents ($p = .0066$), and adults ($p < 1 \times 10^{-5}$) all showed a main effect of reward, whereas adolescents ($p = .0016$) and adults ($p < .0001$), but not children ($p = .65$), showed a reward-by-transition-type interaction effect.

Categorical age groupings reflect rigid but arbitrary delineations between groups. Because developmental changes in learning are likely to be gradual, we tested for age-related differences in the recruitment of each strategy within the full cohort of participants using a continuous age term and its interaction terms (Table 2). The behavioral signatures of both model-free and model-based learning were evident in the full cohort of participants, who showed both a significant main effect of reward (model-free learning; $p < 1 \times 10^{-8}$) and a reward-by-transition-type interaction (model-based learning; $p < 1 \times 10^{-6}$). However, only the model-based learning signature exhibited a significant increase with age (a reward-by-transition-type-by-age interaction, $p = .0004$; Fig. 2b). Analysis of developmental differences in learning strategy in which age was treated as a categorical variable, rather than a continuous variable, yielded similar results (see Supplemental Material). Collectively, these results

Table 1. Logistic Regression Coefficients Indicating the Effects of Previous Reward and Previous Transition Type on First-Stage Choice Repetition Within Each Age Group

Predictor	Effect-size estimate (*SE*)	$\chi^2(1)$	*p*
Children ($n = 20$)			
Intercept	0.61 (0.10)	20.45	$< 1 \times 10^{-5}$
Reward	0.30 (0.08)	11.79	.0006
Transition type	0.01 (0.04)	0.07	.79
Reward × Transition Type	0.02 (0.04)	0.21	.65
Adolescents ($n = 20$)			
Intercept	1.19 (0.23)	17.28	< .0001
Reward	0.22 (0.08)	7.39	.0066
Transition type	0.09 (0.06)	2.34	.13
Reward × Transition Type	0.35 (0.10)	10.00	.0016
Adults ($n = 19$)			
Intercept	1.85 (0.25)	26.32	$< 1 \times 10^{-6}$
Reward	0.56 (0.11)	20.43	$< 1 \times 10^{-5}$
Transition type	0.07 (0.08)	5.06	.024
Reward × Transition Type	0.49 (0.13)	15.64	< .0001

Table 2. Logistic Regression Coefficients Indicating the Effects of Age (as a Continuous Variable), Previous Reward, and Previous Transition Type on First-Stage Choice Repetition for All Participants

Predictor	Effect-size estimate (*SE*)	$\chi^2(1, N = 59)$	*p*
Intercept	1.18 (0.11)	62.05	$< 1 \times 10^{-14}$
Reward	0.34 (0.05)	35.52	$< 1 \times 10^{-8}$
Transition type	0.05 (0.03)	2.29	.130
Age	0.43 (0.11)	13.05	.0003
Reward × Transition Type	0.27 (0.05)	26.00	$< 1 \times 10^{-6}$
Reward × Age	0.08 (0.05)	2.36	.124
Transition × Age	0.01 (0.03)	0.08	.77
Reward × Transition Type × Age	0.18 (0.05)	12.49	.0004

suggest that whereas the behavioral signature of model-free learning was evident across development, the recruitment of model-based learning increased from childhood into adulthood. Finally, the baseline tendency to repeat first-stage choices, independent of previous outcome or transition type, also increased significantly with age (main effect of age, $p < .0003$).

We examined whether participants' choice behavior changed from the first to the second half of the task by repeating the mixed-effects regression analyses with an additional term (half, indicating whether a trial fell in the first or second half of trials) and its interaction terms. There were no significant effects including the half term when age was included as a linear factor (all *p*s > .23) or within each categorical age group (all *p*s > .18), which suggests that age-related differences in learning strategy were stable across the 200 trial sessions.

Logistic regression analysis of choice strategy considers only how the reward and transition type from the previous trial influence subsequent choices. This analytical approach has the advantages of making few assumptions and yielding results that are easily visualized. However, it represents a simple approximation of the choice computations implemented by model-free or model-based reinforcement-learning algorithms, which draw on the full history of choices and outcomes across trials. To verify that our results were consistent across both analytical approaches, we additionally examined the effects of age on the recruitment of model-free and model-based decision making using a computational reinforcement-learning model. The reinforcement-learning model consists of a weighted combination of (a) a model-free temporal-difference algorithm that incrementally updates a fixed value for the first-stage choice based on reward history and (b) a model-based tree-search reinforcement-learning algorithm that evaluates all possible choice options and associated outcomes (Sutton & Barto, 1998). We found that the regression coefficient for the effect of age on the model-based parameter estimate ($\beta_{MB\text{-}age}$) was significantly positive, indicating an increase in the model-based contribution to value computation with age, whereas the coefficient for the age effect on the model-free parameter estimate ($\beta_{MF\text{-}age}$) was not significantly different from 0 (Table 3; see Table S1 in the Supplemental Material). This computational analysis corroborates the age-related increase in the recruitment of model-based strategy that we observed in the logistic regression analysis.

Knowledge of task-transition structure

There were no age-group differences in participants' recollection of the state-transition structure ("Which spaceship traveled to the red planet most of the time?"; $\chi^2(2, N = 43) = 0.6701$, $p = .7153$; 14 of 18 children, 14 of 17 adolescents,

Table 3. Median and 95% Confidence Interval for Parameter Estimates Derived via the Reinforcement-Learning Model Indicating the Effect of Age on Model-Free and Model-Based Evaluation

Parameter	Median estimate	95% CI
β_{MF} (model-free weight)	0.293	[0.214, 0.373]
β_{MB} (model-based weight)	0.314	[0.203, 0.439]
$\beta_{MF\text{-}age}$ (effect of age on model-free weight)	0.058	[–0.021, 0.134]
$\beta_{MB\text{-}age}$ (effect of age on model-based weight)	0.205	[0.088, 0.326]

Note: CI = confidence interval.

and 15 of 17 adults answered correctly; 7 participants were not asked to recall the transition structure). Thus, participants across age groups showed explicit awareness of the transition structure. Next, we examined participants' response times for the second-stage choice as a function of transition type. If participants were not aware of the transition structure, we would expect no response time differences after common transitions compared with rare transitions. A linear mixed-effects model revealed that participants were slower after rare transitions than after common transitions (effect size = 86 ms, SE = 14, Wald $\chi^2(1, 57) = 38.6$, $p < 1 \times 10^{-7}$) and although RTs decreased with age (effect size = −78 ms, $SE = 0.26$, Wald $\chi^2(1, 56.9) = 9.2$, $p = .0036$) there was no transition-type-by-age interaction ($p = .40$; Fig. 3a). These response time effects were unrelated to whether participants repeated the second-stage choice on consecutive trials.

We then tested whether this second-stage slowing—reflecting knowledge of the transition structure—was related to participants' recruitment of a model-based

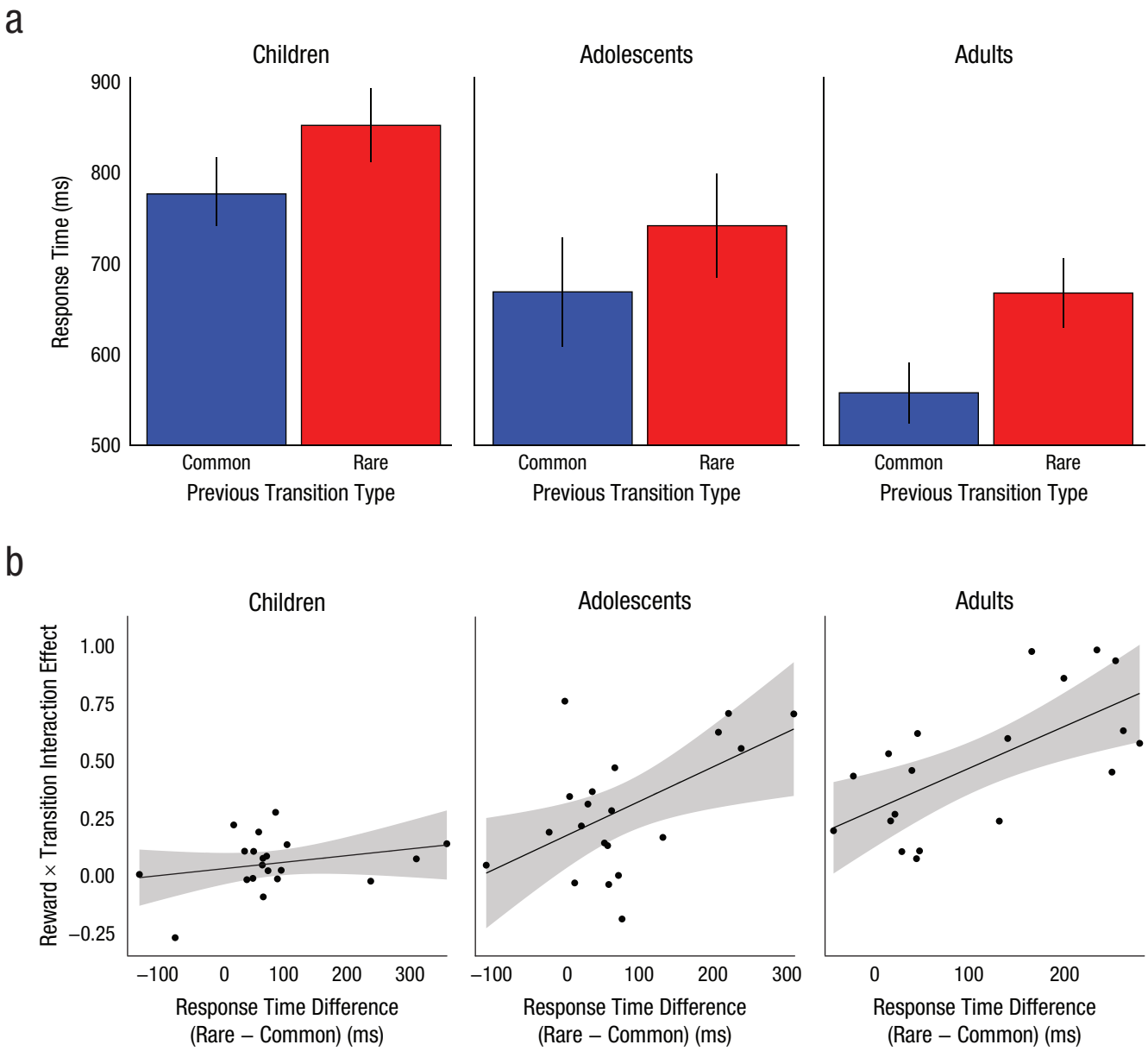

Fig. 3. Second-stage response time results. The bar graphs (a) show response times for choices at the second stage as a function of the preceding transition type for each age group. Error bars represent ±1 *SEM*. The scatterplots (b; with best-fitting regression lines) show the relationship between the reward-by-transition-type interaction effect (the model-based effect estimates) and the difference in response time between choices following rare transitions and those following common transitions for each age group. The gray bands represent ±1 *SEM*.

strategy, as has recently been shown in adults (Deserno, Huys, Boehme, Buchert, & Heinze, 2015). This slowing was associated with model-based choice behavior in adults (r = .66, p = .0023) and adolescents (r = .54, p = .014), but not in children (r = .27, p = .25; Fig. 3b). Collectively, these results suggest that although participants across age groups learned the transition structure of the task, only children did not appear to integrate this information into their first-stage choices.

Discussion

In this study, we examined developmental changes in model-free and model-based evaluation strategies in a sequential decision-making task. We found that children, adolescents, and adults all tended to repeat initial choices that led to rewards, which is the behavioral signature of model-free learning. In contrast, although participants of all ages appeared to distinguish between common and rare transitions, the model-based ability to recruit this knowledge to inform their choices emerged only in the adolescents and continued to strengthen among the adults.

Model-based behavior reflects the ability to use cognitive representations of the environment to inform goal-directed choices. This capacity involves multiple component cognitive processes, including working memory (Otto, Gershman, et al., 2013) and cognitive control (Otto et al., 2015). Executive functions, including working memory (Olesen, Westerberg, & Klingberg, 2004), cognitive control (Munakata, Snyder, & Chatham, 2012), and use of abstract rules or instruction (Bunge & Zelazo, 2006; Decker, Lourenco, Doll, & Hartley, 2015), exhibit a protracted maturational trajectory. For example, in late childhood, children typically transition from a reactive form of cognitive control that supports behavioral correction to a proactive form that involves anticipatory representation of goal-related information (Chatham, Frank, & Munakata, 2009). This ability continues to mature into young adulthood (Braver, 2012). Although gradual development of executive function is widely proposed to confer changes in reward-related behavior (Zelazo & Carlson, 2012), a mechanistic account for this process has been lacking. Our findings suggest that emergent executive functioning may alter reward-guided behavior by providing the necessary foundation for model-based computations. Theoretically, this account obviates any need to invoke a homunculus-like controller that develops an increasing capacity to "override" suboptimal reward-driven responses (Verbruggen, McLaren, & Chambers, 2014). Instead, goal-directed decision making emerges through normative changes in the computations engaged to determine which instrumental actions are optimal. Future studies involving assessment of executive functions might test directly the component cognitive processes that are necessary or sufficient to support the developmental emergence of model-based choice.

Strikingly, although children's choice behavior was strictly model free, they appeared to form a cognitive model of the task. Children, like adolescents and adults, could report the task transitions and exhibited slower response times after rare transitions, providing further evidence of transition structure knowledge. A previous study in adults found that increased magnitude of this slowing predicted greater model-based choice (Deserno et al., 2015). In our study, however, only adolescents and adults showed this correlation. Thus, although children exhibited knowledge of the transition structure, they did not recruit this knowledge prospectively in their subsequent first-stage choices. This emergent model-based ability may reflect a developmental shift from reactive engagement of cognitive control after surprising transitions to proactive cognitive control engaged at the first-stage choice (Braver, 2012; Munakata et al., 2012). Our results accord with a developmental literature describing dissociations between the age at which knowledge is present and the age at which knowledge is behaviorally revealed in task performance (Zelazo, Frye, & Rapus, 1996).

Model-based learning algorithms reproduce several defining features of goal-directed behavior (Daw et al., 2005). Two key properties distinguish goal-directed from habitual behavior: sensitivity to changes in action-outcome contingency and sensitivity to changes in the value of the outcome itself. Perseveration in either condition reveals an action to be habitual (Balleine & O'Doherty, 2009; Dickinson, 1985). In several canonical assays of cognitive development, younger children perseverate with previously rewarded actions after contingency changes. For example, in Piaget's A-not-B task, after an action is reinforced several times (e.g., reach left toward a hidden toy), babies 10 months old or younger are impaired when they must perform a new action (e.g., reach right) in a critical test trial, but by the age of 12 months, this perseveration is no longer seen (Piaget, 1954). In more complex tasks, this developmental emergence of sensitivity to contingency change is observed at later ages (Kirkham, Cruess, & Diamond, 2003; Zelazo et al., 1996). Likewise, sensitivity to outcome devaluation has been reported to emerge across development (Klossek et al., 2008). Model-free learners do not recruit the representations of outcome contingencies and values that are necessary to inform goal-directed behavior. Thus, a parsimonious account may be that these behavioral changes reflect a developmental transition from model-free to model-based action-evaluation processes. This transition may be a general characteristic of cognitive development that occurs at later ages for tasks of greater complexity as the capacity to form and recruit a model of the task improves.

Studies in adult humans and rodents suggest that model-free and model-based learning recruit overlapping but dissociable neural circuits (Balleine & O'Doherty, 2009). Dopaminergic input to the ventral striatum is proposed to carry prediction-error signals that support both model-free and model-based value computations (Daw et al., 2011), in interaction with dissociable dorsal striatal regions (Balleine & O'Doherty, 2009). Model-based learning additionally integrates information about states and outcomes stemming from a relatively more extensive network of regions, including the prefrontal cortex and the hippocampus. By encoding associations between actions and their specific outcomes, the prefrontal cortex may maintain a cognitive model of the task (Balleine & O'Doherty, 2009; Doll et al., 2012). Engagement of the dorsolateral prefrontal cortex during model-based learning (Smittenaar, FitzGerald, Romei, Wright, & Dolan, 2013) may also reflect the contribution of working memory and cognitive control processes (Miller & Cohen, 2001). The hippocampus is widely hypothesized to support model-based learning by encoding sequential relationships between states (Doll et al., 2012; Pennartz, Ito, Verschure, Battaglia, & Robbins, 2011), and hippocampal-striatal connections are proposed to facilitate the integration of state and reward information during choice (Wimmer & Shohamy, 2012).

Developmentally, the striatal prediction-error signals underpinning model-free learning appear relatively mature from childhood onward (Cohen et al., 2010; Van den Bos, Cohen, Kahnt, & Crone, 2012), consistent with our observation of model-free choice behavior across development. In contrast, protracted maturation of corticostriatal connectivity from childhood to adulthood (Somerville & Casey, 2010) may contribute to the gradual development of model-based learning. Collectively, the literature suggests that the developmental emergence of model-based learning may reflect the burgeoning integration of a prefrontal-hippocampal-striatal circuit that recruits learned information about states and outcomes to take goal-directed action (Pennartz et al., 2011; Shohamy & Turk-Browne, 2013). Future studies examining the development of these circuits and their relationship to behavior may elucidate the neurocircuitry underlying the observed developmental increase in model-based learning.

We also observed, beyond developmental changes in reinforcement-learning strategies, an age-related increase in the tendency to repeat a first-stage choice, independent of the preceding transition types and rewards. This residual autocorrelation between successive first-stage choices was low in children, but choices grew "stickier" with age, replicating previous observations of greater response variability in younger individuals (Christakou et al., 2013). Broad sampling of available choice options may reflect a search process that is more strongly biased toward exploration at earlier stages of development (Gopnik, Griffiths, & Lucas, 2015). Such a developmental bias might promote discovery of the environmental structure, as well as the eventual identification of optimal responses.

The balance between model-based and model-free learning may have important implications for real world decision making. Model-free learners may be prone to impulsive choices, repeating actions that previously yielded small immediate rewards and failing to prospectively consider more highly valued long-term goals (Kurth-Nelson, Bickel, & Redish, 2012). Furthermore, insensitivity to outcome devaluation may lead model-free learners to perseverate with previously rewarded actions that are no longer beneficial. In the laboratory, children and adolescents have been found to exhibit greater impulsivity and perseveration in their choices than adults do (Klossek et al., 2008; Mischel et al., 1989), with important real-world repercussions. This may be particularly true during adolescence, when increased exploration and autonomy confers greater opportunity to make new choices along with less parental protection from their consequences. Indeed, the greatest perils of adolescence are those associated with poor choices (e.g., reckless driving, unprotected sex, suicide; Eaton et al., 2006), underscoring the importance of understanding developmental changes in decision making (Hartley & Somerville, 2015). The developmental emergence of model-based learning observed in the present study represents an expansion in the repertoire of evaluative processes available to inform one's actions. This increasing ability to incorporate a model of the complex and changing environment into one's evaluations may promote the maturation of goal-directed decision making from childhood to adulthood.

Action Editor

Brian P. Ackerman served as action editor for this article.

Author Contributions

C. A. Hartley and N. D. Daw developed the study concept. All authors contributed to the study design. J. H. Decker collected the data. J. H. Decker, A. R. Otto, and C. A. Hartley analyzed and interpreted the data. J. H. Decker and C. A. Hartley drafted the manuscript, and N. D. Daw and A. R. Otto provided critical revisions.

Acknowledgments

We thank B. J. Casey for helpful discussions and the participants and families for volunteering their time.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

C. A. Hartley was supported by National Institute on Drug Abuse Grant R03-DA038701, the DeWitt Wallace Reader's Digest Fund, an American Psychological Foundation Esther Katz Rosen Fund grant, and a generous gift from the family of Mortimer D. Sackler. J. H. Decker was supported by National Institute of General Medical Sciences Grant T32-GM007739. N. D. Daw and A. R. Otto were supported by a Scholar Award from the James S. McDonnell Foundation and by National Institute on Drug Abuse Grant 1R01-DA038891.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

Open Practices

All data have been made publicly available via Open Science Framework and can be accessed at https://osf.io/gq7z2. The complete Open Practices Disclosure for this article can be found at http://pss.sagepub.com/content/by/supplemental-data. This article has received the badge for Open Data. More information about the Open Practices badges can be found at https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/ and http://pss.sagepub.com/content/25/1/3.full.

References

Balleine, B. W., & O'Doherty, J. P. (2009). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*, 48–69.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-8) [Software]. Retrieved from http://cran.r-project.org/package=lme4

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*, 106–113. doi:10.1016/j.tics.2011.12.010

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science*, *15*, 118–121. doi:10.1111/j.0963-7214.2006.00419.x

Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences, USA*, *106*, 5529–5533. doi:10.1073/pnas.0810002106

Christakou, A., Gershman, S. J., Niv, Y., Simmons, A., Brammer, M., & Rubia, K. (2013). Neural and psychological maturation of decision-making in adolescence and young adulthood. *Journal of Cognitive Neuroscience*, *25*, 1807–1823. doi:10.1162/jocn_a_00447

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*, 669–671.

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.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*, 1704–1711.

Decker, J. H., Lourenco, F. S., Doll, B. B., & Hartley, C. A. (2015). Experiential reward learning outweighs instruction prior to adulthood. *Cognitive, Affective, & Behavioral Neuroscience*, *15*, 310–320. doi:10.3758/s13415-014-0332-5

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. *Proceedings of the National Academy of Sciences, USA*, *112*, 1595–1600. doi:10.1073/pnas.1417219112

Diamond, A. (2006). The early development of executive functions. In E. Bialystok & F. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 70–95). New York, NY: Oxford University Press.

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *308*, 67–78.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, *22*, 1075–1081. doi:10.1016/j.conb.2012.08.003

Eaton, D. K., Kann, L., Kinchen, S., Ross, J., Hawkins, J., Harris, W. A., . . . Wechsler, H. (2006). Youth risk behavior surveillance—United States, 2005. *Morbidity and Mortality Weekly Report*, *55*, 1–108. doi:10.1016/S0002-9343(97)89440-5

Eppinger, B., Walter, M., Heekeren, H. R., & Li, S.-C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience*, *7*, Article 253. doi:10.3389/fnins.2013.00253

Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science*, *24*, 87–92. doi:10.1177/0963721414556653

Hartley, C. A., & Somerville, L. H. (2015). The neuroscience of adolescent decision-making. *Current Opinion in Behavioral Sciences*, *5*, 101–115. doi:10.1016/j.cobeha.2015.09.004

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, *6*, 449–467. doi:10.1111/1467-7687.00300

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. *Journal of Experimental Psychology: General*, *137*, 39–51. doi:10.1037/0096-3445.137.1.39

Kurth-Nelson, Z., Bickel, W. K., & Redish, A. D. (2012). A theoretical account of cognitive effects in delay discounting. *The European Journal of Neuroscience*, *35*, 1052–1064. doi:10.1111/j.1460-9568.2012.08058.x

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. doi:10.1146/annurev.neuro.24.1.167

Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science*, *244*, 933–938.

Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing cognitive control: Three key transitions. *Current Directions in Psychological Science*, *21*, 71–77. doi:10.1177/0963721412436807

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*, 75–79. doi:10.1038/nn1165

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological Science*, *24*, 751–761. doi:10.1177/0956797612463080

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences, USA*, *110*, 20941–20946. doi:10.1073/pnas.1312011110

Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2015). Cognitive control predicts use of model-based reinforcement learning. *Journal of Cognitive Neuroscience*, *27*, 319–333. doi:10.1162/jocn_a_00709

Pennartz, C. M. A., Ito, R., Verschure, P. F. M. J., Battaglia, F. P., & Robbins, T. W. (2011). The hippocampal-striatal axis in learning, prediction and goal-directed behavior. *Trends in Neurosciences*, *34*, 548–559. doi:10.1016/j.tins.2011.08.001

Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Basic Books.

Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology*, *12*, 427–441. doi:10.1017/S0954579400003096

R Development Core Team. (2015). R: A language and environment for statistical computing (Version 3.2.2) [Computer software]. Retrieved from https://www.r-project.org/index.html

Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology: General*, *142*, 1159–1170. doi:10.1037/a0034461

Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron*, *80*, 914–919. doi:10.1016/j.neuron.2013.08.009

Somerville, L. H., & Casey, B. (2010). Developmental neurobiology of cognitive control and motivational systems. *Current Opinion in Neurobiology*, *20*, 271–277. doi:10.1016/j.conb.2010.01.006

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press. doi:10.1109/TNN.1998.712192

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*, 1247–1255. doi:10.1093/cercor/bhr198

Verbruggen, F., McLaren, I. P. L., & Chambers, C. D. (2014). Banishing the control homunculi in studies of action control and behavior change. *Perspectives on Psychological Science*, *9*, 497–524. doi:10.1177/1745691614526414

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science*, *338*, 270–273. doi:10.1126/science.1223252

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*, 354–360. doi:10.1111/j.1750-8606.2012.00246.x

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63. doi:10.1016/S0885-2014(96)90027-1