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**[2], and **Catherine A. Hartley**[1]

[1]Sackler Institute for Developmental Psychobiology, Weill Cornell Medical College, New York, NY 10065

[2]Center for Neural Science, New York University, New York, NY 10003

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 enables estimation of model-based and model-free contributions to choice. Whereas a model-free strategy was evident in choice behavior across all age groups, evidence of a model-based strategy only emerged during adolescence and continued to increase into adulthood. 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

Introduction

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

Corresponding author: Catherine A. Hartley, PhD, Weill Cornell Medical College, Sackler Institute for Developmental Psychobiology, 1300 York Ave., Box 140, New York, NY 10065, Phone: 212-746-6239, Fax: 212-746-5755, cah2031@med.cornell.edu.

outcomes, with perseveration only decreasing 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 pre-potent 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 our choices in more ecologically relevant behavioral contexts. Recent findings in adults suggest that integrated executive functioning provides a cognitive foundation for more complex decision computations, altering the manner in which reward-related actions are evaluated (Otto, Gershman, Markman, & Daw, 2013; Otto, 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 desire 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 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 upon a given strategy appears 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 reliance upon these alternative forms of learning might change markedly with age. However, the developmental trajectory of these action selection strategies has not yet been examined. Here, 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, model-based influence on choice only emerged in adolescence and continued to increase with age. 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

30 children (aged 8–12), 28 adolescents (13–17), and 22 adults (18–25), completed the task. Based on a previous study (Eppinger, Walter, Heekeren, & Li, 2013) reporting 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 as we expected higher rates of attrition in these age groups. One child and one adolescent were excluded due to inattentiveness (e.g., looking away from the screen, closing eyes) and one adult for complete choice invariance throughout the task. We used two criteria to exclude participants whose behavior was inconsistent with an intention to obtain rewards in the task. First, the proportion of first-stage stay decisions following common transitions had to be at least 0.1 greater following rewarded than unrewarded trials (1 child, 4 adolescents, 1 adult). Second, 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, 4 adolescents – 1 who failed both criteria, 1 adult). These restrictions require that participants appear to be pursuing reward, but are unbiased as to whether they do so via a model-free or model-based strategy. The final sample included 59 participants: 20 children (11 females; 9.80 +− 1.54 years), 20 adolescents (12 females; 15.35 +− 1.39 years), and 19 adults (11 females; 21.63 +−2.03). 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), designed to dissociate model-free and model-based learning strategies, to use a child-friendly narrative and be engaging for a developmental cohort. Prior to 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 concluded with an instruction summary using simple child-friendly terminology. All participants indicated verbally that they understood the instructions prior to starting the task. Participants were tasked with collecting space-treasure (Figure 1A). They made a first-stage choice between two spaceship stimuli, each of which traveled more frequently (70% versus 30%) to one planet than the other. For example, choosing the blue spaceship led to the red planet with 70% probability (the common transition), and led to the purple planet with 30% probability (the rare transition). On each planet, they made a second-stage choice between two stimuli (aliens),

and were rewarded (a picture of 'space treasure') or not (an empty circle) according to a slowly drifting probability (bounded between 0.2 and 0.8). These shifting reward probabilities encouraged participants to explore different choices throughout the task in order to maximize rewards. Participants had 3 seconds to make each choice, followed by a 1 second animation, 1 second of feedback, and a 1 second inter-trial interval. The full game consisted of 200 trials in four blocks separated by breaks.

Critically, this task structure enables dissociation of the relative recruitment of model-free and model-based learning strategies. Whereas a model-based chooser uses a cognitive model of the transitions and outcomes in the task to select actions, a model-free chooser simply repeats previously rewarded actions (Figure 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 is likely to repeat the previous first-stage choice (blue spaceship), regardless of the transition that led to the reward. In contrast, a model-based chooser—taking into account the state transition structure—is likely to switch to the green spaceship, increasing the likelihood of returning to that rewarded state.

Behavioral Analysis

The group logistic regression analysis has been described previously (Daw et al., 2011; Otto, Gershman, et al., 2013). Briefly, a generalized linear mixed-effects regression analysis of group behavior data was performed using the lme4 package for the R-statistics language. First-stage choice (stay/switch from previous trial) was modeled by independent predictors of previous reward (reward/no reward), previous transition (rare/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, and reward-by-transition interaction terms ('random slopes'). The terms of interest are the main effect of reward (model-free term), a reward-by-transition-type interaction effect (model-based term), and the reward-by-age and reward-by-transition-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: child=3.5, adolescent=0.5, adult=0). Additionally, this analysis was performed for each age group separately, removing the age and age-interaction terms. Response time data for the second-stage actions was analyzed similarly using a linear mixed-effects analysis with current transition 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 et al., (2013a), which allows participant's choices to take into account the entire preceding history of rewards. This model 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 falls above (or below) zero, we can conclude that the parameter of interest is significantly

positive (or negative) with 95% confidence. Details of the model fitting procedure are provided in SI Text.

Results

Learning behavior

We assessed participants' recruitment of these alternative 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 previously observed in adults (Daw et al., 2011; Otto, Gershman, et al., 2013). We 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<1e-5) all showed a main effect of reward, whereas only adolescents (p=0.0016) and adults (p<0.0001), but not children (p=0.65), showed a reward-by-transition interaction effect.

Categorical age groupings reflect rigid but arbitrary delineations between groups. As developmental changes in learning are likely to be more 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; p<1e-8) and a reward-by-transition interaction (model-based; p<1e-6). However, only the model-based learning signature exhibited a significant increase with age (reward-by-transition-by-age interaction p=0.0004; Figure 2B). Analysis of developmental differences in learning strategy in which age is treated as a categorical, rather than continuous, variable yields similar results (SI). Collectively, these results suggest that whereas the behavioral signature of model-free learning is evident across development, the recruitment of model-based learning increases 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<0.0003).

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 factor "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-values > 0.23), or within each categorical age group (all p-values > 0.18), suggesting that age-related differences in learning strategy were stable across the 200 trial session.

Logistic regression analysis of choice strategy only considers how the reward and transition 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 the choice computations implemented by

model-free or model-based reinforcement learning algorithms, which draw upon the full history of choices and outcomes across trials. To verify that our results are consistent across both analytical approaches, we additionally examined the effects of age on model-free and model-based decision-making using a computational reinforcement-learning model. We found that the $\beta_{MB\text{-}age}$ term was significantly positive (0.205, 95% CI 0.088–0.326), indicating an increase in the model-based contribution to value computation with age, whereas the $\beta_{MF\text{-}age}$ term (0.058, 95% CI –0.021–0.134) was not significantly different from 0 (Table 3, SI Text, SI Table 1). This computational analysis corroborates the age-related increase in the recruitment of model-based strategy observed in the logistic regression analysis.

Knowledge of 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?") (X-squared = 0.6701, df = 2, p-value = 0.7153, 14/18 children, 14/17 adolescents, and 15/17 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 (RT) for the second-stage choice as a function of transition type. If participants were not aware of the transition structure, we would expect no RT differences following common versus rare transitions. A linear mixed-effects model revealed that participants were slower after rare versus common transitions (86 ms, SE=14, X^2=38.6, p<1e-7), and although RTs decreased with age (–78 ms, SE=0.26, X^2=9.2, p=0.0036), there was no transition-type-by-age interaction (p=0.40) (Figure 3A). These response time effects were unrelated to whether participants repeated the same 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 strategy, as has recently been shown in adults (Deserno, Huys, Boehme, Buchert, & Heinze, 2015). In adults (r=0.66, p=0.0023) and adolescents (r=0.54, p=0.014), this slowing was associated with more model-based choice behavior; however this relationship was not seen in children (r=0.27, p=0.25; Figure 3B). Collectively, these results suggest participants across age groups learned the transition structure of the task, however only children did not appear to integrate this information into their 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, the behavioral signature of model-free learning. In contrast, although participants of all ages appeared to distinguish common from rare transitions, the model-based ability to recruit this knowledge to inform one's choices only emerged in adolescence and continued to strengthen into adulthood.

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 supporting behavioral correction to a proactive form involving 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 greater 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 direct assessment of executive functions might test directly what component cognitive processes 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 structure and exhibited slower response times following rare transitions, providing further evidence of transition knowledge. A previous study in adults found that the magnitude of this slowing predicted greater model-based choice (Deserno et al., 2015). However, in our study, 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 following 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 when it 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 following 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 or younger are impaired in a critical test trial where they must perform a new action (reach right), but by 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). Similarly, 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 computational reward prediction error signals that support both model-free and model-based 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 more extensive network of regions, including the prefrontal cortex (PFC) and the hippocampus. By encoding associations between actions and their specific outcomes, the PFC may maintain a cognitive model of the task (Balleine & O'Doherty, 2009; Doll et al., 2012). Engagement of the dorsolateral PFC during model-based learning (Smittenaar, FitzGerald, Romei, Wright, & Dolan, 2013) may additionally 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 onwards (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, this 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.

Beyond developmental changes in reinforcement learning strategies, we also observed an age-related increase in the tendency to repeat a first-stage choice, independent of the preceding transitions and rewards. This residual autocorrelation between successive first-stage choices was low in children, but choices grew "stickier" into adulthood, 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. Experimentally, children and adolescents have been found to exhibit greater impulsivity and perseveration in their choices than adults (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, 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Balleine BW, O'Doherty JP. Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. Neuropsychopharmacology. 2009; 35(1):48–69. [PubMed: 19776734]

Braver TS. The variable nature of cognitive control: A dual mechanisms framework. Trends in Cognitive Sciences. 2012; 16(2):106–113. DOI: 10.1016/j.tics.2011.12.010 [PubMed: 22245618]

Bunge SA, Zelazo PD. A brain-based account of the development of rule use in childhood. Current Directions in Psychological Science. 2006; 15(3):118–121. DOI: 10.1111/j. 0963-7214.2006.00419.x

Chatham CH, Frank MJ, Munakata Y. Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(14):5529–5533. DOI: 10.1073/pnas.0810002106 [PubMed: 19321427]

Christakou A, Gershman SJ, Niv Y, Simmons A, Brammer M, Rubia K. Neural and psychological maturation of decision-making in adolescence and young adulthood. Journal of Cognitive Neuroscience. 2013; 25(11):1807–1823. DOI: 10.1162/jocn [PubMed: 23859647]

Cohen JR, Asarnow RF, Sabb FW, Bilder RM, Bookheimer SY, Knowlton BJ, Poldrack RA. A unique adolescent response to reward prediction errors. Nature Neuroscience. 2010; 13(6):669–671. [PubMed: 20473290]

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-Based Influences on Humans' Choices and Striatal Prediction Errors. Neuron. 2011; 69(6):1204–1215. [PubMed: 21435563]

Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature Neuroscience. 2005; 8(12):1704–1711. [PubMed: 16286932]

Decker JH, Lourenco FS, Doll BB, Hartley CA. Experiential reward learning outweighs instruction prior to adulthood. Cognitive, Affective & Behavioral Neuroscience. 2015; doi: 10.3758/s13415-014-0332-5

Deserno L, Huys QJM, Boehme R, Buchert R, Heinze H. Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proceedings of the National Academy of Sciences. 2015; 112(5):1595–1600. DOI: 10.1073/pnas.1417219112

Diamond A. The early development of executive functions. Lifespan Cognition: Mechanisms of Change. 2006:70–95.

Dickinson A. Actions and habits: the development of behavioural autonomy. Phiolosophical Transactions of the Royal Society of London. 1985; 308:67–78.

Doll BB, Simon Da, Daw ND. The ubiquity of model-based reinforcement learning. Current Opinion in Neurobiology. 2012; 22(6):1075–81. DOI: 10.1016/j.conb.2012.08.003 [PubMed: 22959354]

Eaton DK, Kann L, Kinchen S, Ross J, Hawkins J, Harris Wa, … Wechsler H. Youth Risk Behavior Surveillance — United States, 2005. Morbidity and Mortality Weekly Report. 2006; 55:1– 108. DOI: 10.1016/S0002-9343(97)89440-5 [PubMed: 16410759]

Eppinger B, Walter M, Heekeren HR, Li SC. Of goals and habits: age-related and individual differences in goal-directed decision-making. Frontiers in Neuroscience. 2013; 7(December): 253.doi: 10.3389/fnins.2013.00253 [PubMed: 24399925]

Gopnik A, Griffiths TL, Lucas CG. When Younger Learners Can Be Better (or at Least More Open-Minded) Than Older Ones. Current Directions in Psychological Science. 2015; 24(2):87–92. DOI: 10.1177/0963721414556653

Hartley CA, Somerville LH. The neuroscience of adolescent decision-making. Current Opinion in Behavioral Sciences. 2015; :1–8. DOI: 10.1016/j.cobeha.2015.09.004

Kirkham NZ, Cruess L, Diamond A. Helping children apply their knowledge to their behavior on a dimension-switching task. Developmental Science. 2003; 6:449–467. DOI: 10.1111/1467-7687.00300

Klossek UMH, Russell J, Dickinson A. The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. Journal of Experimental Psychology. General. 2008; 137(1):39–51. DOI: 10.1037/0096-3445.137.1.39 [PubMed: 18248128]

Kurth-Nelson Z, Bickel WK, Redish AD. A theoretical account of cognitive effects in delay discounting. The European Journal of Neuroscience. 2012; 35(7):1052–64. DOI: 10.1111/j.1460-9568.2012.08058.x [PubMed: 22487035]

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience. 2001; 24:167–202. DOI: 10.1146/annurev.neuro.24.1.167

Mischel W, Shoda Y, Rodriguez MI. Delay of gratification in children. Science (New York, NY). 1989; 244(4907):933–8.

Munakata Y, Snyder HR, Chatham CH. Developing Cognitive Control: Three Key Transitions. Current Directions in Psychological Science. 2012; 21:71–77. DOI: 10.1177/0963721412436807 [PubMed: 22711982]

Olesen PJ, Westerberg H, Klingberg T. Increased prefrontal and parietal activity after training of working memory. Nature Neuroscience. 2004; 7(1):75–79. DOI: 10.1038/nn1165 [PubMed: 14699419]

Otto AR, Gershman SJ, Markman AB, Daw ND. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. Psychological Science. 2013a; 24(5):751–61. DOI: 10.1177/0956797612463080 [PubMed: 23558545]

Otto AR, Raio CM, Chiang A, Phelps Ea, Daw ND. Working-memory capacity protects model-based learning from stress. Proceedings of the National Academy of Sciences of the United States of America. 2013b; 110(52):20941–6. DOI: 10.1073/pnas.1312011110 [PubMed: 24324166]

Otto AR, Skatova A, Madlon-Kay S, Daw ND. Cognitive Control Predicts Use of Model-based Reinforcement Learning. Journal of Cognitive Neuroscience. 2015; 27(2):319–333. DOI: 10.1162/jocn [PubMed: 25170791]

Pennartz CMA, Ito R, Verschure PFMJ, Battaglia FP, Robbins TW. The hippocampal-striatal axis in learning, prediction and goal-directed behavior. Trends in Neurosciences. 2011; 34(10):548–59. DOI: 10.1016/j.tins.2011.08.001 [PubMed: 21889806]

Piaget, J. The construction of reality in the child. New York: Basic Books; 1954.

Posner MI, Rothbart MK. Developing mechanisms of self-regulation. Development and Psychopathology. 2000; 12:427–441. DOI: 10.1017/S0954579400003096 [PubMed: 11014746]

Shohamy D, Turk-Browne NB. Mechanisms for widespread hippocampal involvement in cognition. Journal of Experimental Psychology. General. 2013; 142(4):1159–70. DOI: 10.1037/a0034461 [PubMed: 24246058]

Smittenaar P, FitzGerald THB, Romei V, Wright ND, Dolan RJ. Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. Neuron. 2013; 80(4):914–9. DOI: 10.1016/j.neuron.2013.08.009 [PubMed: 24206669]

Somerville LH, Casey B. Developmental neurobiology of cognitive control and motivational systems. Current Opinion in Neurobiology. 2010; 20(2):271–277. DOI: 10.1016/j.conb.2010.01.006 [PubMed: 20538543]

Van den Bos W, Cohen MX, Kahnt T, Crone Ea. Striatum-medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. Cerebral Cortex. 2012; 22(6):1247–55. DOI: 10.1093/cercor/bhr198 [PubMed: 21817091]

Verbruggen F, McLaren IPL, Chambers CD. Banishing the Control Homunculi in Studies of Action Control and Behavior Change. Perspectives on Psychological Science. 2014; 9(5):497–524. DOI: 10.1177/1745691614526414 [PubMed: 25419227]

Wimmer GE, Shohamy D. Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions. Science. 2012; 338(October):270–273. DOI: 10.1126/science.1223252 [PubMed: 23066083]

Zelazo PD, Carlson SM. Hot and Cool Executive Function in Childhood and Adolescence: Development and Plasticity. Child Development Perspectives. 2012; 6(4) n/a–n/a. doi: 10.1111/j.1750-8606.2012.00246.x

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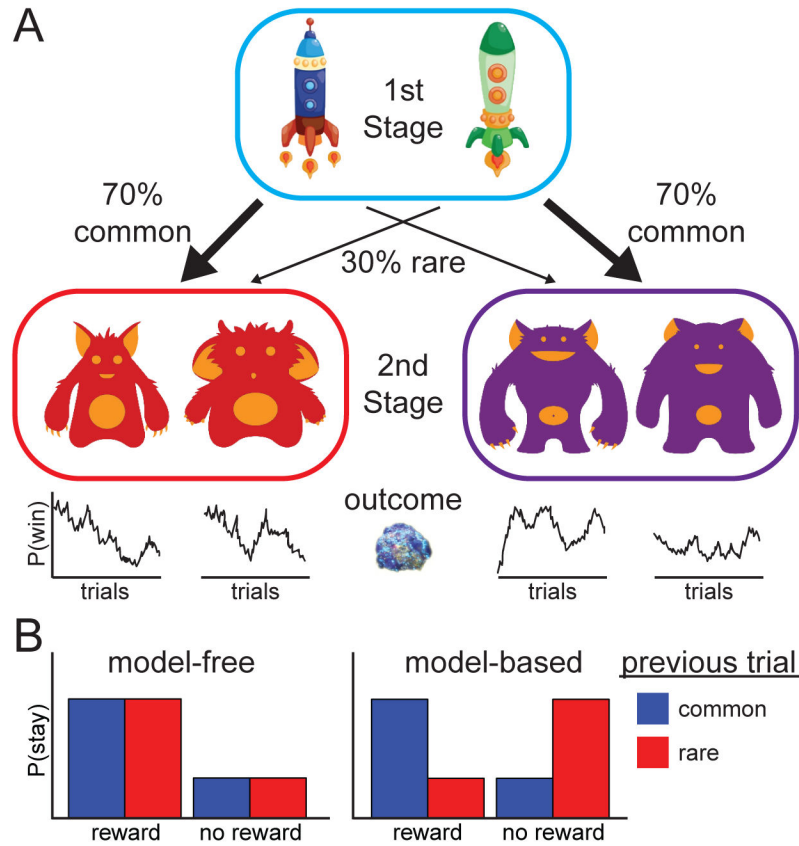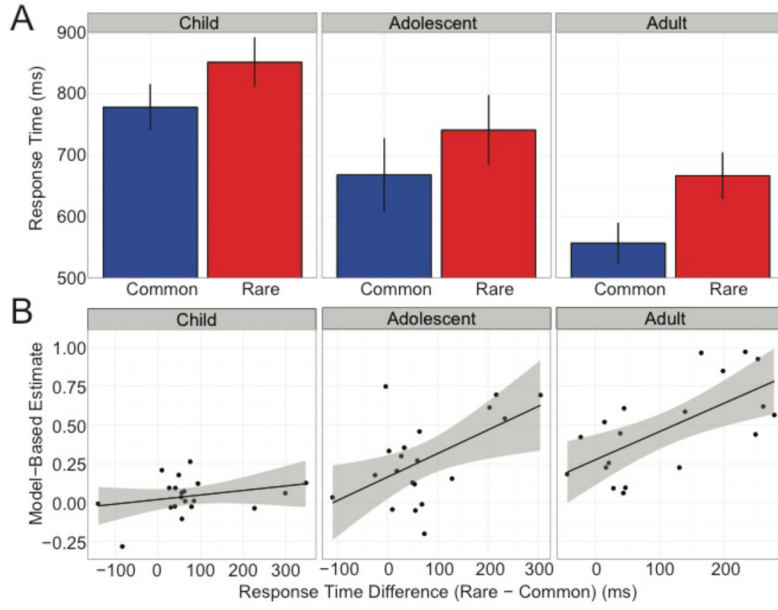
Figure 1.

Sequential "spaceship task" design. (A) For each trial, participants made a first-stage choice between two spaceships, followed by a probabilistic transition to a red or purple planet. At this second stage, participants chose between two aliens and were rewarded with space treasure according to a slowly drifting probability. B) Idealized model-free and model-based behavior. Whereas a model-free chooser relies solely on previous reward to determine whether to repeat a first-stage choice (a main effect of reward), a model-based chooser incorporates both transition and reward information (a reward-by-transition interaction effect).

Figure 2.
(A) Proportion of first-stage stay choices as a function of reward and transition type for each age group. Children showed a model-free strategy, sensitive only to previous reward, whereas adolescents and adults exhibited a mixture of strategies, taking into account both the previous reward and transition type. (B) The model-based signature of behavior (reward-by-transition interaction) was not present in children but increased significantly with age ($p = .0004$).

Figure 3.
A) Children, adolescents, and adults all had slower second-stage response times following a rare transition than a common transition. B) Longer second-stage response times following rare versus common transitions were predictive of more model-based behavior in adolescents and adults, but not in children.

Table 1

Logistic regression results examining the effects of previous reward, and previous transition type on first-stage choice repetition within each age group. Significant p-values (<.05) denoted in bold.

Predictor	Estimate (SE)	X2 (df=1)	p-value
Child (N=20)			
Intercept	0.61 (0.10)	20.45	**<1e-5**
Reward	0.30 (0.08)	11.79	**0.0006**
Transition	0.01 (0.04)	0.07	0.79
Reward by Transition	0.02 (0.04)	0.21	0.65
Adolescent (N=20)			
Intercept	1.19 (0.23)	17.28	**<0.0001**
Reward	0.22 (0.08)	7.39	**0.0066**
Transition	0.09 (0.06)	2.34	0.13
Reward by Transition	0.35 (0.10)	10.00	**0.0016**
Adult (N=19)			
Intercept	1.85 (0.25)	26.32	**<1e-6**
Reward	0.56 (0.11)	20.43	**<1e-5**
Transition	0.07 (0.08)	5.06	**0.024**
Reward by Transition	0.49 (0.13)	15.64	**<0.0001**

Table 2

Logistic regression results examining the effects of continuous age, previous reward, and previous transition type on first-stage choice repetition. Significant p-values (<.05) denoted in bold.

Predictor	Estimate (SE)	X2 (df=1)	p-value
Full Group (N=59)			
Intercept	1.18 (0.11)	62.05	**<1e-14**
Reward	0.34 (0.05)	35.52	**<1e-8**
Transition	0.05 (0.03)	2.29	0.130
Age	0.43 (0.11)	13.05	**0.0003**
Reward by Transition	0.27 (0.05)	26.00	**<1e-6**
Reward by Age	0.08 (0.05)	2.36	0.124
Transition by Age	0.01 (0.03)	0.08	0.77
Reward by Transition by Age	0.18 (0.05)	12.49	**0.0004**

Table 3

Median and 95% confidence interval for parameters relating age to model-free and model-based evaluation.

Parameter	Description	Median	Lower 95%	Upper 95%
β_{MF}	Model-free weight	0.293	0.214	0.373
β_{MB}	Model-based weight	0.314	0.203	0.439
$\beta_{MF\text{-}age}$	Age on model-free weight	0.058	−0.021	0.134
$\beta_{MB\text{-}age}$	Age on model-based weight	0.205	0.088	0.326