Developmental Cognitive Neuroscience 25 (2017) 272–280

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: http://www.elsevier.com/locate/dcn

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.

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; Klossek 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.

http://dx.doi.org/10.1016/j.dcn.2016.10.005

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 lambda (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 alpha (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