Neuron
Report

A Reinforcement Learning Mechanism Responsible for the Valuation of Free Choice

Jeffrey Cockburn,[1] Anne G.E. Collins,[1] and Michael J. Frank[1,*]
[1]Department of Cognitive, Linguistic and Psychological Sciences; Brown Institute for Brain Science, Brown University, Providence, RI 02912, USA
*Correspondence: michael_frank@brown.edu
http://dx.doi.org/10.1016/j.neuron.2014.06.035

SUMMARY

Humans exhibit a preference for options they have freely chosen over equally valued options they have not; however, the neural mechanism that drives this bias and its functional significance have yet to be identified. Here, we propose a model in which choice biases arise due to amplified positive reward prediction errors associated with free choice. Using a novel variant of a probabilistic learning task, we show that choice biases are selective to options that are predominantly associated with positive outcomes. A polymorphism in DARPP-32, a gene linked to dopaminergic striatal plasticity and individual differences in reinforcement learning, was found to predict the effect of choice as a function of value. We propose that these choice biases are the behavioral byproduct of a credit assignment mechanism responsible for ensuring the effective delivery of dopaminergic reinforcement learning signals broadcast to the striatum.

INTRODUCTION

An organism's fitness is determined by its ability to avoid hazard while in pursuit of reward (Orr, 2009). In light of this, choice is a terrifically advantageous faculty as it offers a handhold through which an organism can manipulate the environment in terms of its needs. However, the advantages of choice come at a cost. The cognitive overhead associated with identifying needs, opportunities, candidate actions, and selecting among them implies that choice-governed behavior will be more demanding than simple stimulus-driven response. Indeed, evidence suggests that complex choices can be aversive (Iyengar and Lepper, 2000). Nevertheless, humans and animals alike demonstrate a preference for choice (Bown et al., 2003; Leotti and Delgado, 2011, 2014) and for options that were freely chosen over equally valued options that were not (Egan et al., 2007; Lieberman et al., 2001; Sharot et al., 2009, 2010).

Preference for freely chosen options has been viewed through the lens of cognitive dissonance theory, whereby the psychological tension that comes with having to choose among equally valued options is resolved postchoice by reevaluating those options in favor of what was chosen (Festinger, 1962). Tversky (1972) has argued along similar reevaluative lines but suggests that the process of choosing alters the importance ascribed to option features and, as such, postchoice valuation takes place in a different context where feature weights favor the chosen option. More recently, studies have shown that humans not only prefer options they have already chosen but also exhibit a bias if given the option of making a choice or not (Bown et al., 2003). Striatal blood-oxygen-level-dependent (BOLD) signal has been found to correlate with both change in option valuation postchoice (Sharot et al., 2009) and with the preference for choice (Leotti and Delgado, 2011, 2014). However, the neural mechanisms through which these biases emerge have been left unexplained and so too have their functional significance. Here, we ask whether choice biases might be diagnostic of a more general adaptive mechanism.

We aimed to determine whether a computational mechanism summarizing reinforcement learning (RL) processes in the basal ganglia (BG) could explain these findings. We hypothesized that free-choice biases are the behavioral byproduct of a feedback loop involving the BG and the midbrain dopamine (DA) system, a mechanism through which positive reward prediction errors (RPEs) encoded by DA cells are preferentially amplified following free choice (see Figure 2A). We propose that this feedback loop alleviates a credit assignment problem in the brain by providing a channel through which dopaminergic learning signals come to preferentially target the BG whenever it has taken part in the agent's endogenous action selection process that yielded a positive outcome.

Our hypothesis was motivated by three key findings. First, exogenously driven behavior is controlled cortically, whereas endogenous choice-driven behavior depends on additional recruitment of the BG (Brown and Marsden, 1998; François-Brosseau et al., 2009). Second, BOLD signal change in human striatum is correlated with both the anticipation of choice (Leotti and Delgado, 2011, 2014) and preference for freely chosen options (Sharot et al., 2009). Third, striatal, but not frontal, DA was found to increase as a function of choice in rodents (St Onge et al., 2012). Together, these findings suggest that choice engages the BG and influences striatal DA levels.

Anatomical work points to a mechanism through which the BG could modulate dopaminergic signals. Tonically active cells in the substantia nigra pars reticulata (SNr) send inhibitory projections onto DA cells of the substantia nigra pars compacta (SNc) (Joel and Weiner, 2000). A decrease in SNr activity (as occurs when an action is gated through the BG) reduces the SNr's inhibitory influence over the SNc, thus facilitating DA release into the striatum (Lee et al., 2004). In other words, the SNr applies

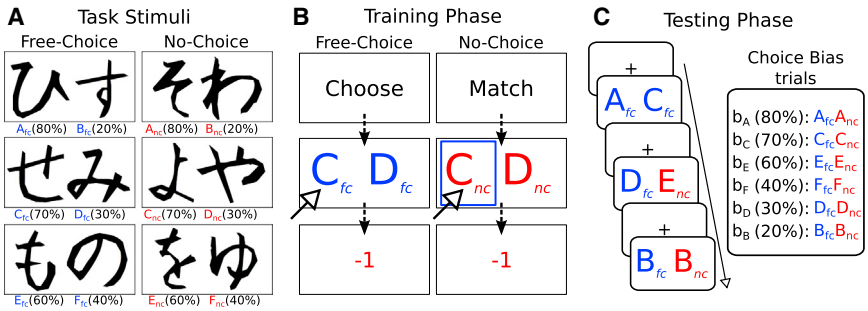

Figure 1. Experimental Task Design

(A) Example free-choice (fc) and no-choice (nc) stimuli used in the task with associated reward probabilities shown. (B) Training phase: one stimulus pair is presented per trial. Participants are asked to select one of the two available options. Participants were alerted to the free-choice (Choose) or no-choice (Match) condition prior to stimulus presentation. On free-choice trials, participants were free to choose either option, but on no-choice trials, participants were forced to select the framed stimulus. Probabilistic feedback followed option selection. (C) Test phase: participants were repeatedly asked to choose the best option among all possible option pairings. Participants were free to choose either stimulus on all trials, but no feedback was provided. Choice bias was quantified according to performance on trials where equally rewarded free-choice and no-choice options were paired.

a break on SNc activity. This break is released when the BG gates an action, thereby increasing the upper range of DA release into the striatum should DA cells be driven to burst by additional afferent SNc inputs.

A biophysical model of these structures has demonstrated that striatal activity associated with action selection inhibits the SNr, which in turn disinhibits SNc cells and thereby increases phasic DA bursting (Lobb et al., 2011). Furthermore, incorporating such a mechanism into a biologically constrained model of the BG has been shown to increase learning signal fidelity and improve performance in complex environments (O'Reilly and Frank, 2006).

In line with these observations, we hypothesized that phasic DA bursts are preferentially amplified when they are associated with BG-gated actions. As such, gated actions should develop inflated values relative to actions that were not, which would emerge behaviorally as a preference for freely chosen options. This mechanism implies that choice bias magnitudes should be determined by RPE history; and as such, we aimed to systematically assess biases across a range of option values and RPE histories. If choice bias is governed by dopaminergic learning in the BG, we also reasoned that genetic variation of dopaminergic striatal plasticity and reward learning should be predictive of individual choice bias differences. Specifically, we focused on the DARPP-32 gene, a gene that has been linked to reward learning and individual differences in learning to pursue (as opposed to avoid) options (Doll et al., 2011; Frank et al., 2007, 2009; Stipanovich et al., 2008).

We tested our hypothesis by administering a novel variant of a probabilistic learning task previously shown to be sensitive to striatal function across a range of conditions (Doll et al., 2011; Frank et al., 2004, 2007) and also allowed for a direct comparison between preference for options that were freely chosen relative to those that were not. Participants were asked to sample and learn about six pairs of stimuli of various expected values (see Figure 1A), with probabilistic feedback (either a point gained or lost) awarded after each selection (see Figure 1B). Participants were randomly presented with one of the six stimulus pairs on each training trial: three of those stimulus pairs allowed participants to choose freely between both options (fc: free-choice), whereas the other three stimulus pairs forced participants to pick a preselected stimulus (nc: no-choice). Critically, no-choice trials were yoked to free-choice trials to ensure identical sampling and reward feedback across conditions.

Following the training phase, a test phase probed what had been learned. Participants were presented with all possible option pairings and asked to select the better of the two on each trial (see Figure 1C). Here, participants were free to choose on all trials but were no longer given feedback. Importantly, to isolate the value of choice across a range of reward probabilities, participants encountered trials where they had to choose between free-choice and no-choice options with identical reward contingencies.

We formalized the behavioral implications of our hypothesis using a computational model of striatal RL. To better represent the BG's anatomical structure, we extended the standard actor-critic architecture, which has been suggested to formalize some of the BG's core functionality (O'Doherty et al., 2004), by including opponent actor weights that contribute positive ("Go") and negative ("NoGo") evidence for each option. These distinct sets of action weights embody the functional implications of D_1- and D_2-expressing striatal medium spiny neurons that take part in the direct and indirect pathways, respectively (Frank, 2005). In this model, RPEs are proportionally added to Go weights according to learning rate parameter α_g, while simultaneously having an opposing subtractive effect on NoGo weights according to learning rate parameter α_n. Thus, this extended actor comprises an opponent process where Go and NoGo weights come to represent positive and negative outcome expectancy, respectively, and where choice probability is a function of the relative difference between Go and NoGo weights for each action under consideration. This opponent actor model captures a wide range of data associated with striatal dopamine manipulations on learning and incentive motivation that cannot be captured by standard single actor models (Collins and Frank, 2014). Here, we further investigated the impact of free choice amplification of positive prediction errors in this framework (see Supplemental Information available online for model details).

RESULTS

To investigate the behavioral consequences of our hypothesis, we augmented the core BG model to include a parameter, α_{fc+},