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

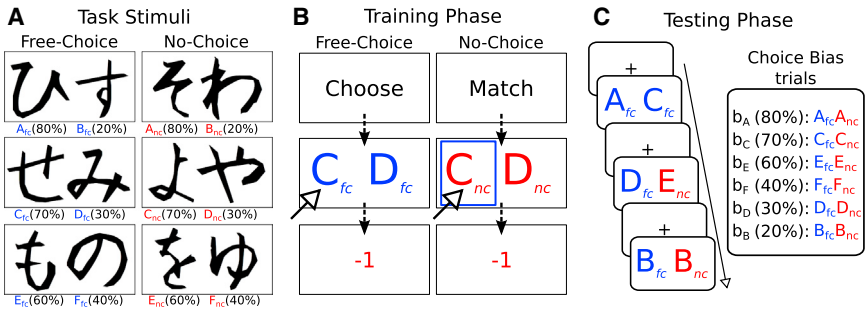

Figure 1. Experimental Task Design

(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+},

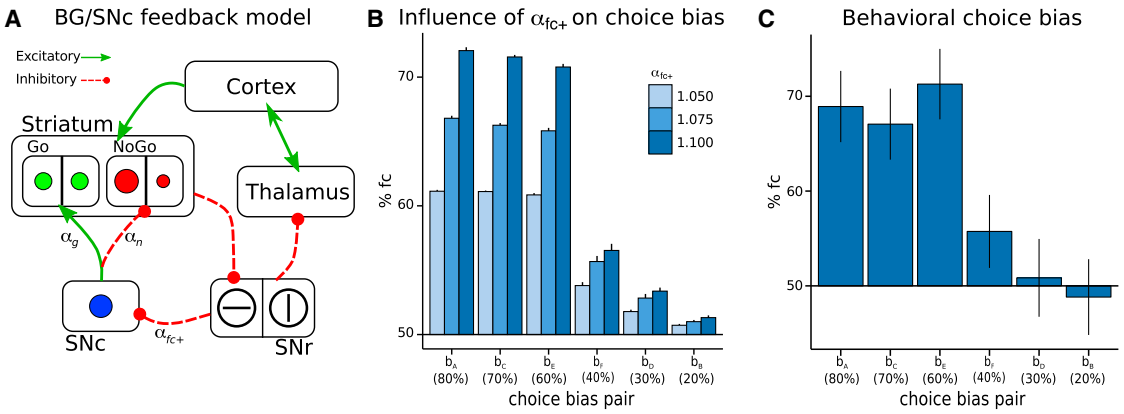

Figure 2. Positive RPE Amplification Mechanism and Choice Bias Patterns

(A) A simplified diagram of the BG/SNc feedback circuitry. Sensory and motor information is projected to the BG via corticostriatal projections, where it is channeled through both the direct Go (green circles) and indirect NoGo (red circles) pathways, providing positive and negative evidence for each action, respectively, before converging at the substantia nigra pars reticulata (SNr). The activity pattern depicted here illustrates a case of balanced Go activity for two candidate actions, but differential NoGo activity, leading to gating of the right-most action. Vertical bar indicates the gated action to the thalamus. The same disinhibitory mechanism that gates thalamocortical actions also disinhibit SNc dopaminergic signals via SNr-SNc projections, thereby allowing reinforcement signals to be amplified when the BG gates an action. The degree of free-choice amplification due to this mechanism is captured by α_{fc+}. (B) Model generated choice bias for a range of α_{fc+} values as a function of reward contingency, computed as the percentage of trials where the free-choice (fc) option was selected. (C) Participant preferences on choice bias trials as a function of reward contingency, calculated as the percentage of choice bias trials where the free-choice (fc) option was selected. Error bars indicate SEM.

which modulated the influence of positive free-choice RPEs on both Go and NoGo weights. We then exposed the model to the experimental task while systematically varying α_{fc+}. Figure 2B illustrates the effect α_{fc+} has on preferences for free-choice options over equally valued no-choice options. When RPEs are balanced across choice conditions ($\alpha_{fc+} = 1$), free-choice and no-choice options share identical RPE histories, and as such, the model exhibits no choice bias whatsoever. However, as α_{fc+} increases it plays a larger role in shaping the action weights, particularly for rewarding free-choice options that are associated with positive RPEs more often than not, resulting in a widening preference for rewarding free-choice options. Human performance mirrored the model's response pattern (Figure 2C). Participants exhibited a strong preference for rewarding free-choice options over their no-choice counterparts ($z = 6.84$, $p < 0.001$), but showed no such preference for nonrewarding options ($z = 0.71$, $p = 0.48$).

Before probing the choice bias in more detail, we first establish that preferences are consistent across the various options by leveraging the behavioral choice bias pattern to infer a relational option value structure (see Figure 3A). Here, no-choice values take on the true expected value of each option (e.g., $nc_{80\%} = E[A_{nc}]$), whereas free-choice values are adjusted according to the behaviorally quantified choice biases for each option (e.g., $fc_{80\%} = E[A_{fc}] + b_A$). The structure depicted in Figure 3A can then be tested by comparing preferences for any given option over any of the others.

The value added due to free choice leads to a discrepancy between equally rewarded options (e.g., $b_A = fc_{80\%} - nc_{80\%}$). This discrepancy should translate to a consistent free-choice preference modulation across all other options (e.g., $fc_{80\%} - fc_{30\%} = (nc_{80\%} + b_A) - fc_{30\%}$, and $fc_{80\%} - nc_{60\%} = (nc_{80\%} + b_A) - nc_{60\%}$). We probed for this predicted pattern by assessing accuracy on trials involving the most rewarding free-choice and no-choice options, entering root option (A_{fc}, A_{nc}), and paired option (C_{fc}, E_{fc}, … D_{nc}) as factors in a logistic regression (see Figure 3B). This analysis revealed an overall A_{fc} performance gain that was consistent across all paired options (main effect of root option: $\chi^2(1) = 29.23$, $p < 0.01$; main effect of paired option: $\chi^2(7) = 138.02$, $p < 0.01$; interaction: $\chi^2(7) = 9.25$, $p > 0.2$). Adjusting A_{fc} trial accuracy by the behaviorally quantified choice bias (Figure 3B: $A_{fc} - b_A$) rendered performance indistinguishable from A_{nc} trials, indicating that A_{fc} performance benefits were consistent with the choice bias across all option pairings (main effect of root: $\chi^2(1) = 0.15$, $p > 0.6$; main effect of pairing: $\chi^2(7) = 127.43$, $p < 0.01$; interaction: $\chi^2(7) = 9.26$, $p > 0.2$). The expected preference patterns were also observed across pairs involving the worst options (see Figure 3C).

In summary, participant behavior was consistent with the value structure depicted in Figure 3A across a range of independent option pairings (see Figure S2 for a more complete analysis). These results demonstrate that participants learned the relative values of both free-choice and no-choice options, that preferences were internally consistent across stimulus pairs, and, as predicted by our computational model, that choice bias effects are more pronounced across rewarding options.

Impact of Reward Probability on Choice Amplification of Option Values

The effect of valence on choice bias patterns appears categorical: values are boosted for positive but not negative options, but with no further modulation of value according to reward probabilities. However, the model predicts that reward probability shapes action weights but with opposing effects on Go versus

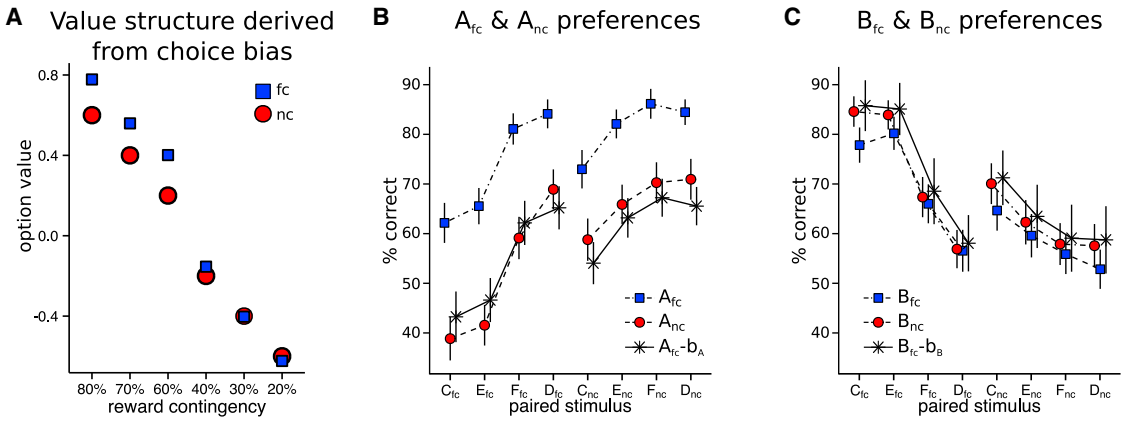

Figure 3. Derived Value Structure and Implied Preference Patterns

(A) The option value structure derived from the empirically quantified choice bias. No-choice options (nc) take on true expected values. Free-choice options (fc) take on the true expected values adjusted according to the choice bias for each option. (B) Percent correct (choice of more rewarding option) across trials involving A_{fc} or A_{nc}. (C) Percent correct across trials involving B_{fc} or B_{nc}. All error bars represent SEM.

NoGo weights. As illustrated in Figure 4A, amplified positive RPEs have a greater impact on Go weights for more rewarding options (e.g., A_{fc}), where positive RPEs are more frequently encountered. This increases the model's preference to choose more rewarding free-choice options, which in itself would drive greater choice biases with increasing reward probability (i.e., $b_A > b_C > b_E$). However, this is counteracted by the opposite pattern in NoGo weights, which are larger for more moderately rewarding options (e.g., E_{fc}). Here, amplified positive RPEs act to disproportionately decrease NoGo weights for these less rewarding options. This decreases the model's preference to avoid moderately rewarding free-choice options, which on its own would drive greater choice biases with decreasing reward probability (i.e., $b_A < b_C < b_E$).

The opposing biases that develop across Go/NoGo weights give rise to a balanced effect of choice across rewarding options when Go/NoGo learning is symmetrical (see Figure 2B). However, effects of choice in each pathway can be exposed when Go/NoGo learning is asymmetrical, as captured by the relative balance between α_g and α_n learning rate parameters. As illustrated in Figure 4B (and see Figure S4), when Go learning is emphasized ($\alpha_g > \alpha_n$), the Go pathway's choice bias dominates, resulting in a bias that is strongest for the most rewarding option, and decreases parametrically according to the probability of reward ($b_A > b_C > b_E$). The opposite choice bias pattern ($b_A < b_C < b_E$) expressed by the NoGo pathway emerges when NoGo learning is emphasized ($\alpha_g < \alpha_n$). Thus, the computational model predicts that choice bias patterns should vary as a function of learning asymmetries and individual differences thereof.

We sought to determine whether the behavioral consequences of Go/NoGo learning asymmetries were consistent with the model-generated choice bias patterns. To do so, we analyzed behavior according to DARPP-32 genotype, a gene associated with striatal dopamine function (Stipanovich et al., 2008), and asymmetries in Go versus NoGo learning (Doll et al., 2011; Frank et al., 2007, 2009). First, by fitting model parameters to the trial-by-trial behavioral data, we established that DARPP-32 genotype was associated with identifiable Go/NoGo learning asymmetries. Bayesian model selection (Stephan et al., 2009) demonstrated that TT-carriers were best fit by a model that enforced an $\alpha_g > \alpha_n$ learning rate asymmetry, whereas C-carriers were best fit by a model that enforced an $\alpha_g < \alpha_n$ learning rate asymmetry (see supplemental procedures and Table S2 for model fitting and comparison). In line with the model's prediction, and as illustrated in Figure 4C, analyses revealed a gene group by value interaction ($\chi^2(2) = 9.88$, $p = 0.007$). Analysis within each gene group in isolation revealed that C-carriers ($\alpha_g < \alpha_n$) exhibited a $b_A < b_C < b_E$ choice bias pattern ($z = 2.85$, $p < 0.005$), whereas TT-carriers ($\alpha_g > \alpha_n$) exhibited the reverse $b_A > b_C > b_E$ choice bias pattern ($z = -1.83$, $p = 0.068$).

DISCUSSION

Consistent with our hypothesis that choice selectively amplifies positive RPEs, free choice biases were observed across rewarding but not nonrewarding options. We also show evidence suggesting that amplified positive RPEs have differential effects depending on the relative balance between learning from positive and negative outcomes. The implications of this model for choice bias are such that RPE amplification increases Go weights and decreases NoGo weights for rewarding options, simultaneously increasing the propensity to choose the most strongly rewarded options (e.g., A_{fc}) and reducing the propensity to avoid moderately rewarding options (e.g., E_{fc}). As seen in our sample as a whole, a balanced choice bias pattern emerges across rewarding options when learning is balanced across Go/NoGo pathways. This supports prior work linking choice biases to BG function (Leotti and Delgado, 2011, 2014; Sharot et al., 2009), and extends those findings by providing a mechanistic explanation supported by quantitative behavioral and modeling evidence. This mechanism also provides a natural explanation for the boundary conditions under which choice

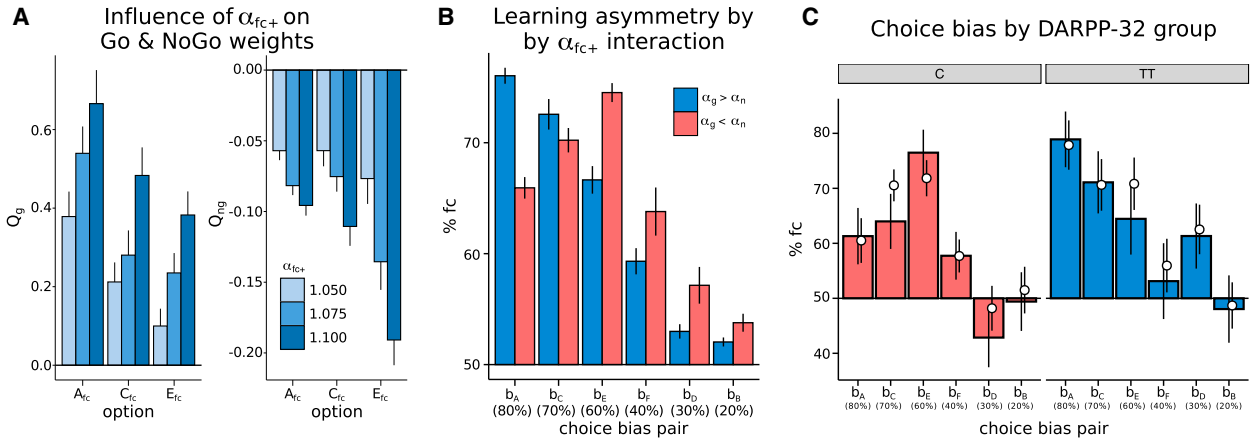

Figure 4. Effects of Positive RPE Amplification on Actor Weights and Its Interaction with Learning Asymmetries

(A) The effect of amplified positive RPEs on Go (Q_g) and NoGo (Q_{ng}) weights. Go weights for the most rewarding options are preferentially amplified, increasing the model's propensity to select those options in accordance with the degree of amplification ($A_{fc} > C_{fc} > E_{fc}$). NoGo weights for the least rewarding options are preferentially dampened, decreasing the model's propensity to avoid those options in accordance with the degree of dampening ($A_{fc} < C_{fc} < E_{fc}$). (B) The interaction between α_{fc+} and the α_g / α_n asymmetry. (C) Choice-bias according to DARPP-32 gene groups (C or TT) as a function of expected value. Bars represent behavioral data, and points represent options preferences recovered from the best fitting model. Error bars indicate SEM.

bias is observed, whereby options associated with more positive prediction errors exhibit a greater free choice bias.

Our results also demonstrate that the relative balance of learning in the opponent pathways determines the degree to which amplified positive RPEs accumulated in Go or NoGo weights, yielding distinct choice bias patterns. We found that DARPP-32 genotype, a gene variant that has been linked to striatal plasticity and asymmetries in learning from positive versus negative RPEs (Doll et al., 2011; Frank et al., 2007, 2009; Stipanovich et al., 2008), predicted individual choice bias differences. This result not only informs us of individual differences in their own right, but more generally, it exposes the underlying mechanism of choice bias rooted in the BG's circuitry, and is particularly diagnostic of our model. Importantly, the choice bias patterns observed across DARPP-32 gene groups argues against an attentional explanation of choice bias, wherein a choice bias emerges because engagement is greater during endogenous action selection. Indeed, evidence suggests that engagement is greater when being rewarded, which often leads to a confound between reward and attention (Maunsell, 2004). However, DARPP-32 C-carriers show a weaker bias for more reliably rewarded options, which is consistent with our computational model, but contrary to the predicted effects of reward on task engagement.

Similar patterns of choice biases have been reported, with more pronounced biases for selected relative to rejected options (Sharot et al., 2009), and stronger biases for options predictive of gains relative to those predictive of losses (Leotti and Delgado, 2014). However, as reported by Leotti and Delgado (2014), choice biases for aversive options are subject to both contextual effects and high variability. Indeed, our sample included a small number of participants (n = 16 of 80 total) that exhibited a bias for aversive options. However, these biases were unsystematic, with individual participants exhibiting both a preference and aversion for different negative options. Furthermore, we could identify neither genetic, nor computational, nor behavioral predictors of negative option choice biases, suggesting that mechanisms beyond dopaminergic striatal learning play a role in shaping biases for negative options.

We have focused our efforts on investigating the interaction between choice and learning. However, humans also exhibit a preference for choice in general (Bown et al., 2003; Leotti and Delgado, 2011, 2014), an issue we have not tackled here. This choice preference may reflect the inherent value of choice, but it may also reflect a learned benefit for the general state of choice. As alluded to previously, freely chosen outcomes are more likely to meet an organism's needs, and as such, an organism could learn to favor environmental states that afford choice as better predictors of reward. Choice may also come to be favored via temporal difference learning, whereby augmented option values, amplified via the BG/SNc mechanism discussed here, are propagated to option predictive states. Although these possibilities offer interesting avenues for future research, they all appear to be at odds with reports of choice aversion (Iyengar and Lepper, 2000). We suggest that choice may be rendered appetitive or aversive according to the degree of choice conflict driven by candidate options. Complex choice spaces, such as those employed by Iyengar and Lepper (2000), could potentially generate a sufficiently high degree of choice conflict so as to prohibit option selection, perhaps via inhibitory mechanisms such as the subthalamic nucleus (Frank, 2006).

Although our results suggest that choice is associated with better learning from positive RPEs, it raises the obvious question of why this should be the case. The BG is commonly thought to embody a gating function that biases action selection (Ashby et al., 2007; Frank, 2005; Mink, 1996). This gating function is embodied by the connectivity of medium spiny neurons in the dorsal striatum, which take part in either the direct Go or the indirect NoGo pathway (Alexander and Crutcher, 1990). The relative difference between Go and NoGo activity for candidate

actions proposed by cortical-striatal projection determines which action will be gated through to the thalamus, providing a selection bias for candidate actions (Frank, 2005). Phasic DA signals from the SNc are thought to provide the learning signals required to develop appropriate Go and NoGo associations via downstream effects on D_1 and D_2 receptors.

However, action selection is not determined by the BG alone, and as such, executed actions may differ from actions preferred by the BG. Thus, broadcasting RL signals uniformly across the brain presents a credit assignment problem: how do the circuits involved ensure that reinforcement is reliably delivered to the neural systems coding for the action that was actually executed? Solutions to this problem often invoke the notion that only recently active neurons will be subject to DA-modulated plasticity (Schultz, 2002; Wickens et al., 1996). However, this allows for reinforcement in systems engaged by the decision-making process, but whose actions were not ultimately executed. The problem is compounded further within BG itself, where cells coding for actions that were considered but not ultimately gated could be inappropriately shaped by dopaminergic signals (see Figure 2A).

One solution to the BG's credit assignment problem is provided if DA neurons in the SNc are themselves gated specifically when the BG gates an action, a mechanism that could be embodied by disinhibitory projections from the SNr. According to this scheme, the BG helps solve its own credit assignment problem by providing the SNc with information diagnostic of action gating. This signal primes DA cells in the SNc such that phasic DA bursts broadcast to the striatum will be more effective whenever the BG takes part in the action selection process. Pushing this idea further, the SNr could potentially provide the SNc with information that not only signals action gating, but a richer signal diagnostic of the action itself. This information could then be integrated by the SNc so as to structure phasic DA signals in a way that preferentially targets populations of striatal cells encoding the gated action. Although it is not currently known whether the SNc's projection architecture is capable of supporting such a richly structured signal, we believe this to be a computationally alluring possibility.

The problem of credit assignment is often overlooked: somehow, the brain's learning signals are delivered to the correct addresses across a labyrinthine landscape. Recent work has proposed that learning signals are decomposed into effector-specific components when appropriate (Gershman et al., 2009), suggesting that learning signals can indeed be structured. We have proposed a relatively simple mechanism through which learning signals may be endowed with such structure and have demonstrated that this mechanism explains why organisms prefer options they have freely chosen. In short, learning signals associated with freely chosen options are more efficacious owing to the engagement of a feedback loop between the BG and DA systems tasked with mitigating the challenge of credit assignment, which emerges behaviorally as a free-choice bias.

EXPERIMENTAL PROCEDURES

Sample

Eighty participants were recruited from Brown University and the Providence, Rhode Island community. Six participants did not demonstrate task learning and were excluded from the analysis (quantified as below chance performance on trials involving A_{fc} or B_{fc}). However, the main results reported here hold when all participants are included in the analysis. The Brown University Human Research Committee approved all task procedures.

Participants provided a saliva sample following the task. We obtained genotype data on an SNP of the *PPP1R1B* (DARPP-32) gene (rs907094), and an SNP of the *DRD2* gene (rs6277), both of which have been associated with striatal DA function (Hirvonen et al., 2009; Stipanovich et al., 2008), and the val158met SNP of the *COMT* gene (rs4680), which has been associated with extracellular DA levels in prefrontal cortex (Huotari et al., 2002; Matsumoto et al., 2003). DARPP-32 allele frequency was 7:35:32 (C/C:C/T:T/T), *DRD2* allele frequency was 21:45:8 (C/C:C/T:T/T), and *COMT* allele frequency was 13:38:23 (MetMet:ValMet:ValVal). All SNPs were in Hardy–Weinberg equilibrium (χ^2 values < 1, p values > 0.4). Categorical gene groups were defined by grouping the most infrequent homogeneous allele carriers with heterogeneous allele carriers, producing DARPP-32 C:TT groups (42:32), *DRD2* CC:T (21:53) groups, and *COMT* Met:ValVal groups (51:23). There was a positive correlation between DARPP-32 and *DRD2* T allele frequency ($r(72) = 0.26$, $p = 0.03$), and a trend for a positive correlation between categorical gene groups ($r(72) = 0.19$, $p = 0.1$). We controlled for this interaction by including *DRD2* as a covariate in all statistical models investigating genetic predictors of behavior. There were no correlations between *COMT* and either DARPP-32 or *DRD2* allele frequency or gene groups (all p values > 0.4).

Most participants self-identified as Caucasian (49 participants). To control for population stratification as a potential confounding factor, we included race as a covariate in all statistical models that also included gene group as a factor. However, the results reported in the main paper hold when minority groups were removed from the analysis.

Procedures

Participants viewed pairs of visual stimuli that are not easily verbalized. During the training phase, six different stimulus pairs were presented in random order, with probabilistic feedback following option selection (either a point gained or lost). Choosing option A_{fc} led to positive feedback 80% of the time, whereas choosing option B_{fc} led to positive feedback only 20% of the time. $C_{fc}D_{fc}$ and $E_{fc}F_{fc}$ pairs were less reliable (see Figure 1A for all reward contingencies).

On free-choice trials participants could choose either option presented to them. No-choice trials were yoked to free-choice trials to ensure identical sampling and reinforcement histories between conditions. The selected option and feedback from each free-choice trial was recorded and used to generate a yoked no-choice trial. For example, if C_{fc} was selected on a $C_{fc}D_{fc}$ trial, and -1 was provided as feedback, a corresponding $C_{nc}D_{nc}$ trial would be generated that forced the selection of C_{nc} (indicated by a blue frame surrounding that option) and provide -1 as feedback. Thus, options in both conditions were sampled the same number of times and delivered the same feedback.

Participants completed at least four and at most six training blocks. Each consisted of 20 exposures to each of the six option pairs. A performance criterion evaluated at the end of each block ensured that all participants were at approximately same performance level before advancing to the test phase (65% selection of A_{fc}, 60% selection of C_{fc}, 50% selection of E_{fc}). Participants could advance to the test phase of the task after completing a minimum of four blocks and exceeding the practice criterion or after six blocks.

Participants were subsequently tested on a full permutation of all possible option pairings (eight pairings of each choice bias pair, and four repetitions of all other pairings) in random order. Participants were free to choose either option on each test trial but were no longer provided feedback (see Supplemental Experimental Procedures for a detailed description of the experimental design).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2014.06.035.

Accepted: June 27, 2014
Published: July 24, 2014

REFERENCES

Alexander, G.E., and Crutcher, M.D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci. *13*, 266–271.

Ashby, F.G., Ennis, J.M., and Spiering, B.J. (2007). A neurobiological theory of automaticity in perceptual categorization. Psychol. Rev. *114*, 632–656.

Bown, N.J., Read, D., and Summers, B. (2003). The lure of choice. J. Behav. Decis. Making *16*, 297–308.

Brown, P., and Marsden, C.D. (1998). What do the basal ganglia do? Lancet *351*, 1801–1804.

Collins, A.G.E., and Frank, M.J. (2014). Opponent Actor Learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol. Rev. *121*, 337–366.

Doll, B.B., Hutchison, K.E., and Frank, M.J. (2011). Dopaminergic genes predict individual differences in susceptibility to confirmation bias. J. Neurosci. *31*, 6188–6198.

Egan, L.C., Santos, L.R., and Bloom, P. (2007). The origins of cognitive dissonance: evidence from children and monkeys. Psychol. Sci. *18*, 978–983.

Festinger, L. (1962). A Theory of Cognitive Dissonance. (Stanford: Stanford University Press).

François-Brosseau, F.-E., Martinu, K., Strafella, A.P., Petrides, M., Simard, F., and Monchi, O. (2009). Basal ganglia and frontal involvement in self-generated and externally-triggered finger movements in the dominant and non-dominant hand. Eur. J. Neurosci. *29*, 1277–1286.

Frank, M.J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J. Cogn. Neurosci. *17*, 51–72.

Frank, M.J. (2006). Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. Neural Netw. *19*, 1120–1136.

Frank, M.J., Seeberger, L.C., and O'reilly, R.C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science *306*, 1940–1943.

Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T., and Hutchison, K.E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc. Natl. Acad. Sci. USA *104*, 16311–16316.

Frank, M.J., Doll, B.B., Oas-Terpstra, J., and Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat. Neurosci. *12*, 1062–1068.

Gershman, S.J., Pesaran, B., and Daw, N.D. (2009). Human reinforcement learning subdivides structured action spaces by learning effector-specific values. J. Neurosci. *29*, 13524–13531.

Hirvonen, M.M., Laakso, A., Någren, K., Rinne, J.O., Pohjalainen, T., and Hietala, J. (2009). C957T polymorphism of dopamine D2 receptor gene affects striatal DRD2 in vivo availability by changing the receptor affinity. Synapse *63*, 907–912.

Huotari, M., Gogos, J.A., Karayiorgou, M., Koponen, O., Forsberg, M., Raasmaja, A., Hyttinen, J., and Männistö, P.T. (2002). Brain catecholamine metabolism in catechol-O-methyltransferase (COMT)-deficient mice. Eur. J. Neurosci. *15*, 246–256.

Iyengar, S.S., and Lepper, M.R. (2000). When choice is demotivating: can one desire too much of a good thing? J. Pers. Soc. Psychol. *79*, 995–1006.

Joel, D., and Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. Neuroscience *96*, 451–474.

Lee, C.R., Abercrombie, E.D., and Tepper, J.M. (2004). Pallidal control of substantia nigra dopaminergic neuron firing pattern and its relation to extracellular neostriatal dopamine levels. Neuroscience *129*, 481–489.

Leotti, L.A., and Delgado, M.R. (2011). The inherent reward of choice. Psychol. Sci. *22*, 1310–1318.

Leotti, L.A., and Delgado, M.R. (2014). The value of exercising control over monetary gains and losses. Psychol. Sci. *25*, 596–604.

Lieberman, M.D., Ochsner, K.N., Gilbert, D.T., and Schacter, D.L. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. Psychol. Sci. *12*, 135–140.

Lobb, C.J., Troyer, T.W., Wilson, C.J., and Paladini, C.a. (2011). Disinhibition bursting of dopaminergic neurons. Front. Syst. Neurosci. *5*, 1–8.

Matsumoto, M., Weickert, C.S., Akil, M., Lipska, B.K., Hyde, T.M., Herman, M.M., Kleinman, J.E., and Weinberger, D.R. (2003). Catechol O-methyltransferase mRNA expression in human and rat brain: evidence for a role in cortical neuronal function. Neuroscience *116*, 127–137.

Maunsell, J.H.R. (2004). Neuronal representations of cognitive state: reward or attention? Trends Cogn. Sci. *8*, 261–265.

Mink, J.W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. Prog. Neurobiol. *50*, 381–425.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science *304*, 452–454.

O'Reilly, R.C., and Frank, M.J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput. *18*, 283–328.

Orr, H.A. (2009). Fitness and its role in evolutionary genetics. Nat. Rev. Genet. *10*, 531–539.

Schultz, W. (2002). Getting formal with dopamine and reward. Neuron *36*, 241–263.

Sharot, T., De Martino, B., and Dolan, R.J. (2009). How choice reveals and shapes expected hedonic outcome. J. Neurosci. *29*, 3760–3765.

Sharot, T., Velasquez, C.M., and Dolan, R.J. (2010). Do decisions shape preference? Evidence from blind choice. Psychol. Sci. *21*, 1231–1235.

St Onge, J.R., Ahn, S., Phillips, A.G., and Floresco, S.B. (2012). Dynamic fluctuations in dopamine efflux in the prefrontal cortex and nucleus accumbens during risk-based decision making. J. Neurosci. *32*, 16880–16891.

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., and Friston, K.J. (2009). Bayesian model selection for group studies. Neuroimage *46*, 1004–1017.

Stipanovich, A., Valjent, E., Matamales, M., Nishi, A., Ahn, J.-H.H., Maroteaux, M., Bertran-Gonzalez, J., Brami-Cherrier, K., Enslen, H., Corbillé, A.-G.G., et al. (2008). A phosphatase cascade by which rewarding stimuli control nucleosomal response. Nature *453*, 879–884.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychol. Rev. *79*, 281.

Wickens, J.R., Begg, A.J., and Arbuthnott, G.W. (1996). Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. Neuroscience *70*, 1–5.