Neurobiology of Learning and Memory 123 (2015) 84–91

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

journal homepage: www.elsevier.com/locate/ynlme

A frontal dopamine system for reflective exploratory behavior

Nathaniel J. Blanco [a,*], Bradley C. Love [b], Jessica A. Cooper [a], John E. McGeary [c,d], Valerie S. Knopik [c,d], W. Todd Maddox [a]

[a] Department of Psychology, University of Texas at Austin, 78712, United States
[b] Experimental Psychology, University College London, WC1H OAP, UK
[c] Department of Psychiatry & Human Behavior, Warren Alpert Medical School, Brown University, 02903, United States
[d] Division of Behavioral Genetics, Rhode Island Hospital, 02903, United States

ARTICLE INFO

Article history:
Received 6 April 2015
Revised 10 May 2015
Accepted 15 May 2015
Available online 22 May 2015

Keywords:
Exploratory behavior
Decision-making
Genetics
COMT
Computational modeling
Dopamine

ABSTRACT

The *COMT* gene modulates dopamine levels in prefrontal cortex with Met allele carriers having lower COMT enzyme activity and, therefore, higher dopamine levels compared to Val/Val homozygotes. Concordantly, Val/Val homozygotes tend to perform worse and display increased (interpreted as inefficient) frontal activation in certain cognitive tasks. In a sample of 209 participants, we test the hypothesis that Met carriers will be advantaged in a decision-making task that demands sequencing exploratory and exploitive choices to minimize uncertainty about the reward structure in the environment. Previous work suggests that optimal performance depends on limited cognitive resources supported by prefrontal systems. If so, Met carriers should outperform Val/Val homozygotes, particularly under dual-task conditions that tax limited cognitive resources. In accord with these a priori predictions, Met carriers were more resilient in the face of cognitive load, continuing to explore in a sophisticated manner. We fit computational models that embody sophisticated reflective and simple reflexive strategies to further evaluate participants' exploration behavior. The Ideal Actor model reflectively updates beliefs and plans ahead, taking into account the information gained by each choice and making choices that maximize long-term payoffs. In contrast, the Naïve Reinforcement Learning (RL) model instantiates the reflexive account of choice, in which the values of actions are based only on the rewards experienced so far. Its beliefs are updated reflexively in response to observed changes in rewards. Converging with standard analyses, Met carriers were best characterized by the Ideal Actor model, whereas Val/Val homozygotes were best characterized by the Naive RL model, particularly under dual-task conditions.

1. Introduction

Effective decision-making requires a balance of exploratory and exploitative behavior (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Cohen, McClure, & Yu, 2007; Hills et al., 2015). For example, consider the problem of choosing the best route to work. Routes change over time because of construction, changes in traffic patterns, etc. such that one cannot be certain which route is currently best. In this non-stationary environment, one either chooses the best-experienced route so far (i.e., exploit) or tries a route that was inferior in the past but now may be superior (i.e., explore). Which actions a commuter should take in a series of choices is a non-trivial problem as optimal decision-making requires factoring in uncertainty about the state of the environment. An actor who excessively exploits will fail to notice when another action becomes superior. Conversely, an actor who excessively explores incurs an opportunity cost by frequently forgoing the high-payoff option.

Our focus is on the timing of exploratory choices. People should explore when they are uncertain about the state of the environment. *Reflective* belief-updates do this by incorporating predictions about unobserved changes in the environment. For example, a reflective belief-updater would increase their belief that an inferior route has improved as more time passes since the last observation because it becomes more likely that disruptive construction will have completed. In contrast, a *reflexive* belief-updater is only informed by direct observations of rewards and, therefore, does not fully utilize environmental structure to update beliefs and guide actions resulting in randomly timed exploratory choices.

This distinction closely echoes contemporary dual-system Reinforcement Learning (RL) approaches in which a reflexive, computationally parsimonious model-free controller competes for

* Corresponding author at: Department of Psychology, University of Texas at Austin, 108 E. Dean Keeton Stop A8000, Austin, TX 78712-1043, United States.
E-mail address: nathanblanco@gmail.com (N.J. Blanco).

http://dx.doi.org/10.1016/j.nlm.2015.05.004
1074-7427/

control of behavior with a reflective, model-based controller situated in prefrontal cortex (Daw, Niv, & Dayan, 2005). Previous work on exploration and exploitation indicates that *reflective* choice is resource intensive, perhaps relying on prefrontal systems (Badre, Doll, Long, & Frank, 2012; Otto, Knox, Markman, & Love, 2014). Correspondingly, populations that have reduced executive function, such as those experiencing depressive symptoms, are impaired in reflective decision making (Blanco, Otto, Maddox, Beevers, & Love, 2013), as are individuals under a secondary task load that exhausts limited cognitive resources (Otto et al., 2014).

Here, we test the hypothesis that reflective exploration is mediated by prefrontal systems by examining differences in the functional Val158Met polymorphism within the *COMT* gene (rs4680). The *COMT* gene modulates dopamine levels in prefrontal cortex with Met allele carriers having lower COMT enzyme activity and, therefore, higher dopamine levels, compared to Val/Val homozygotes (Gogos et al., 1998; Yavich, Forsberg, Karayiorgou, Gogos, & Mannisto, 2007; Kaenmaki et al., 2010). Val/Val homozygotes tend to perform worse on executive tasks and display increased frontal activation that may reflect inefficient processing compared to Met-carriers (Blasi et al., 2005; Winterer et al., 2006; Tan et al., 2007). Animal studies examining set-shifting behavior also indicate the crucial role of PFC dopamine (Stefani & Moghaddam, 2006), which can be manipulated by COMT (Tunbridge, Bannerman, Sharp, & Harrison, 2004). In humans, the *COMT* genotype predicts participants' ability to adapt behavior on a trial-by-trial basis (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007), has been associated with performance on reversal learning tasks (Nolan, Bilder, Lachman, & Volavka, 2004), and has been linked to uncertainty-based exploration (Frank, Doll, Oas-Terpstra, & Moreno, 2009). But, the influence of the Val158Met polymorphism on cognitive function is debated, with some conflicting results. A recent meta-analysis concluded that there was little or no association between *COMT* genotype and scores on a set of standard cognitive tests (e.g. the Wisconsin Card Sorting task), though a reliable association was found between Met/Met genotype and higher IQ (Barnett, Scoriels, & Munafò, 2008).

It may be that *COMT* genotype has a more specific or subtle influence on cognition than is measured by many of the standard behavioral tests. Here we directly assessed the role of *COMT* variation in an exploratory decision-making task. We use computational models, related to reflective and reflexive exploration, to provide a clearer picture of the behavioral data. The main prediction is that Met carriers will explore reflectively, whereas Val/Val homozygotes will rely on simpler reflexive strategies.

One possibility is that the additional dopamine available for Met carriers functions more as a reserve rather than to facilitate cognitive function in general. We predict that Met carriers will be more resilient when cognitive resources are taxed under dual-task load.

2. Materials and methods

We examined associations of *COMT* variants with exploratory strategies by using a paradigm termed the "Leapfrog" task (Knox, Otto, Stone, & Love, 2012), a variant of the "bandit" task (Sutton & Barto, 1998) that is specifically designed to evaluate exploratory behavior. In this task (Fig. 1), one of two options provides a higher reward than the other. With a fixed probability on each trial, the currently inferior option can increase in value, becoming the better option. Because the relative superiority of the options switches over time, participants must choose between *exploiting* the option with the highest observed reward and *exploring* to see whether the other option has surpassed it. This task is ideally suited to evaluate

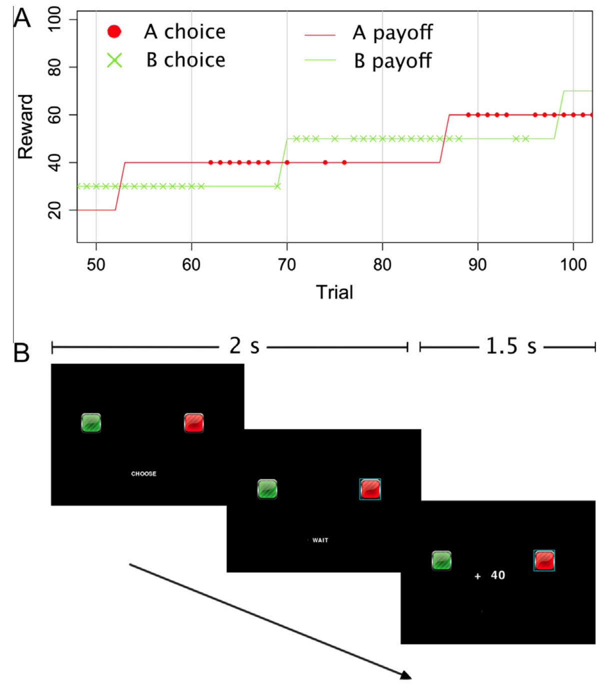

Fig. 1. The Leapfrog task: example choices over 100 trials. On any trial the lower option might, with a probability of 0.075, increase its reward by 20 points, surpassing the other option. The relative superiority of the two options alternates as their reward values "leapfrog" over one another. The lines represent the true reward values, the dots a participant's choices.

the timing of exploratory choices and to what extent they are guided by uncertainty in the environment, distinguishing reflective from reflexive choice strategies.

To tax mechanisms that support reflective exploration, which are thought to be resource intensive, participants in the dual-task condition also performed a tone counting task. Dual-task manipulations using tone counting are known to increase the prevalence of reflexive exploration strategies (Otto et al., 2014). More generally, secondary tasks that exhaust working memory resources tend to increase reliance on implicit strategies (Foerde, Knowlton, & Poldrack, 2006; Zeithamova & Maddox, 2006) and cognitively inexpensive model-free choice strategies (Otto, Gershman, Markman, & Daw, 2013; Gershman, Markman, & Otto, 2014).

2.1. Models evaluated

We fit computational models that embody reflective and reflexive strategies to participants' data to evaluate their exploration behavior. The *Ideal Actor* model reflectively updates beliefs and plans ahead, taking into account the information gained by each choice and making choices that maximize long-term payoffs. Action-values are a product of both expected rewards and the potential to reduce uncertainty about the state of the environment. In contrast, the *Naïve RL* model instantiates the reflexive account of choice, in which the values of actions are based only on the rewards experienced so far. Its beliefs are updated reflexively in response to observed changes in rewards.

Both models incorporate a Softmax choice rule (Sutton & Barto, 1998), which chooses options as a function of the computed action-values. The Softmax inverse temperature is a free parameter in both models. Critically, the action-values used in the Softmax choice rule differ between the two models, leading to qualitative differences in exploratory behavior. The Naïve RL model explores with equal probability on every trial, whereas the probability of