Behavioural Brain Research 306 (2016) 52–63

Contents lists available at ScienceDirect

Behavioural Brain Research

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

Research report

The role of the medial prefrontal cortex in updating reward value and avoiding perseveration

Laskowski C.S.[a], Williams R.J.[b], Martens K.M.[c], Gruber A.J.[a], Fisher K.G.[a], Euston D.R.[a,*]

[a] Department of Neuroscience, Canadian Centre for Behavioural Neuroscience, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada
[b] Faculty of Health Science, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada
[c] Department of Pathology and Laboratory Medicine, School of Medicine, University of British Columbia, 2211 Wesbrook Mall, Vancouver, British Columbia V6T 2B5, Canada

HIGHLIGHTS

- Value updating in rats was studied using 3 choices with occasionally shuffled reward amounts.
- Lesions of medial prefrontal cortex did not affect the propensity to explore.
- Lesioned animals switched more slowly and obtained less reward overall.
- Most rats preferred certain maze arms, a tendency exacerbated by lesions.
- Strong place preference was correlated with dorsal medial prefrontal damage.

ARTICLE INFO

Article history:
Received 2 October 2015
Received in revised form 8 February 2016
Accepted 3 March 2016
Available online 7 March 2016

Keywords:
Reward
Value
Prefrontal cortex
Decision making
Exploration
Reinforcement learning

ABSTRACT

The medial prefrontal cortex (mPFC) plays a major role in goal-directed behaviours, but it is unclear whether it plays a role in breaking away from a high-value reward in order to explore for better options. To address this question, we designed a novel 3-arm Bandit Task in which rats were required to choose one of three potential reward arms, each of which was associated with a different amount of food reward and time-out punishment. After a variable number of choice trials the reward locations were shuffled and animals had to disengage from the now devalued arm and explore the other options in order to optimise payout. Lesion and control groups' behaviours on the task were then analysed by fitting data with a reinforcement learning model. As expected, lesioned animals obtained less reward overall due to an inability to flexibly adapt their behaviours after a change in reward location. However, modelling results showed that lesioned animals were no more likely to explore than control animals. We also discovered that all animals showed a strong preference for certain maze arms, at the expense of reward. This tendency was exacerbated in the lesioned animals, with the strongest effects seen in a subset of animals with damage to dorsal mPFC. The results confirm a role for mPFC in goal-directed behaviours but suggest that rats rely on other areas to resolve the explore-exploit dilemma.

1. Introduction

Both human and animal studies have suggested that impaired functioning of the medial prefrontal cortex (mPFC) leads to difficulty navigating the many complex decisions that life presents. Bechara and colleagues developed a risk-reward decision task, the Iowa Gambling Task, which is diagnostic for the decision deficits underlying mPFC dysfunction [1]. Human neuroimaging studies using this task have indicated that, in healthy adults, ventromedial PFC (vmPFC) activity correlates positively with optimal decisions [2] while those with decision-making impairments, such as gambling addiction, show reduced vmPFC activity in this task compared to controls [3]. Consistent with this view, numerous studies in rats have shown impairments in risk-reward decisions after mPFC disruption [4–8]. However, the specific decision-making deficit associated with mPFC damage remains unclear. Studies in rodents have implicated mPFC in a wide range of decision-related processes including delayed alternation [9,10], memory for task rules [11–14], memory retrieval [15–17,for a review see 18], task switch-

* Corresponding author.
E-mail addresses: david.euston@gmail.com, euston@uleth.ca (D.R. Euston).

http://dx.doi.org/10.1016/j.bbr.2016.03.007
0166-4328/

ing [19,20] and – in more dorsal aspects – weighing effort and reward [21–23].

One reason the mPFC may be critical to a wide range of decision tasks is its key role in representing the predicted outcome of different response options. Single-cell recording studies in both rodents and non-human primates show unambiguously that mPFC represents value. For example, a range of studies in both monkeys and rodents have shown that regions of mPFC represent action selection values, expected reward values, and reward outcome values [24–30]. These findings are complemented by numerous behavioural studies showing that animals with mPFC dysfunction are impaired in value-based decisions, specifically, in avoiding responses leading to devalued outcomes [31–33].

The mPFC is also known to play a critical role in behavioural flexibility [e.g.,19,20,34,35]. Such flexibility might be of ecological relevance in the context of foraging, where animals have to constantly make trade-offs between exploiting a currently discovered food resource and exploring for potentially richer alternatives [36]. Failure to exploit results, naturally, in less food intake while failure to explore results in missed opportunities. It has been proposed that activation of noradrenergic projections from the locus coeruleus enables exploratory behaviour [37]. The locus coeruleus, in turn, receives its primary cortical input from the mPFC, suggesting that this region may play an integral role in modulating the explore-exploit balance [37,38]. Consistent with this hypothesis, the dynamics of neural ensemble firing in the dorsal mPFC (i.e., the anterior cingulate cortex, ACC) show distinct changes when animals switch from exploration to exploitation [39]. A recent human functional magnetic resonance imaging (fMRI) study found enhanced activity in frontopolar cortex correlated with explore decisions [42]. Another fMRI study found specific activation in a region of dorsal ACC when subjects decided to forgo immediate reward to select a new set of choice stimuli [40]. While the functionally equivalent regions in rodents remains unclear, it is plausible that at least one is located in the rat mPFC [41]. To date, the role of rodent mPFC in explore-exploit decisions has not been explicitly tested in a behavioural paradigm.

To test the role of the mPFC in explore-exploit decisions and further explore the role of rat mPFC in value-based decisions, we adopted the *n*-armed Bandit Task used by Daw et al. [42] to study decision-making in humans, for use with rats. The term "1-arm Bandit" is a humorous reference to old-style slot machines. In Daw et al.'s version of the task, subjects chose one of four bandits, each offering different, probabilistic reward values, in order to determine which offered the best reward. Over time the mean amount of reward offered by each bandit changed gradually so that a bandit with a high pay-off initially would gradually decrease in value and vice-versa. In this task a person choosing among several options must balance the impulse to select the bandit they believe offers the highest reward (an exploit decision) against the opportunity to gather information about the value of other options which have the potential to be more valuable (an explore decision).

One of the strengths of the *n*-armed Bandit Task is that behaviour can be parametrically fit using reinforcement learning models, yielding insights about the underlying decision processes [43,44]. In these models, the decision maker (referred to as an "agent") constructs an internal representation of the value of each of the available response options. The core idea of reinforcement learning is that the value of these options is updated dynamically based on the difference between expected and actual reward received. In other words, value updates are driven by the reward prediction error. These algorithms gained considerable credence after it was discovered that the dopaminergic neurons in the ventral tegmental area, in fact, carry such a reward prediction signal [45]. While the equations are provided in the methods section, we briefly describe the parameters available from the reinforcement learning model. The speed with which the model updates internal values based on the reward prediction error is quantified via a learning rate parameter, α. High α values mean the subject adapts very rapidly to changing reward values and pays attention to only the last few outcomes whereas low α values indicate that the animal integrates outcome information over many trials.

Reinforcement learning models also have a second stage, referred to as the decision rule, by which the internal value estimates are translated into a choice. The Daw et al. study used a "softmax" rule, which says that the choice probability is monotonically related to the value of that choice. The model contains a parameter, β, which is homologous to inverse temperature, and quantifies the randomness of the choices. High β values mean the subject is strongly biased to the choice with the highest predicted reward. In the extreme, as β becomes very large, the algorithm implements a "winner-take-all" strategy. Low β values, on the other hand lead to more random choices. In the extreme, all choices are equally distributed, independent of expected value. If an exploratory decision is conceptualized as a choice of any bandit other than the one yielding the highest predicted outcome, then low β means more exploratory decisions. The reinforcement learning hence allows us to quantify both the speed of learning new values and the balance of exploration/exploitation.

We designed a novel version of the *n*-armed Bandit Task for rats with three arms. The task required an adaptive response to changing reward amounts on each of three arms of a radial maze. As with the original Daw et al. version of the task, the task allows us to examine both the speed with which animals adjust to changing reward amounts and the degree of exploration, using a reinforcement learning model to quantify the results. We hypothesized that rats with lesions centred on the prelimbic (PL) region of the mPFC would have deficits in updating value and would hence be slow to adjust when reward outcomes changed. Further, we hypothesized they would have deficits in the exploration/exploitation balance. In particular, we expected they would perseverate on the current high reward arm and, hence, explore other options less frequently.

2. Materials and methods

2.1. Subjects

Subjects were male Long–Evans rats ($n = 26$; Charles River Laboratories, Senneville, QC). Four animals in the lesion group were excluded from analysis, one due to poor performance during pre-training (i.e., the rat repeatedly jumped off the elevated maze), and three due to inadequate or unilateral lesions. Animals weighed 380–450 g and were 3.5–4.5 months old (mean = 131 days) at the start of the experiment. Animals were singly housed in a temperature-controlled colony room under a 12 h reverse light cycle (lights off at 10:00 AM.). All experiments were performed in accordance with the ordinances set by the Canadian Council of Animal Care, and experimental protocols were approved by the University of Lethbridge Animal Welfare Committee.

2.2. Surgery

Subjects were randomly assigned to receive either bilateral lesions of the prelimbic region of the mPFC (n = 13) or sham surgeries (n = 13). Animals were injected with buprenorphine (0.03 mg/kg; Sigma Alderich, Oakville, ON), 30 min prior to being anesthetized with 1–3% isoflurane (2-chloro-2-(difluoromethoxy)-1,1,1-trifluoro-ethane, Abbott Laboratories, Abbott Park, IL) and then secured in a stereotaxic frame. After reaching a deep anesthetic plane, a craniotomy was performed and a 33 gauge stainless steel injection needle attached to a 5 μl Hamilton syringe