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

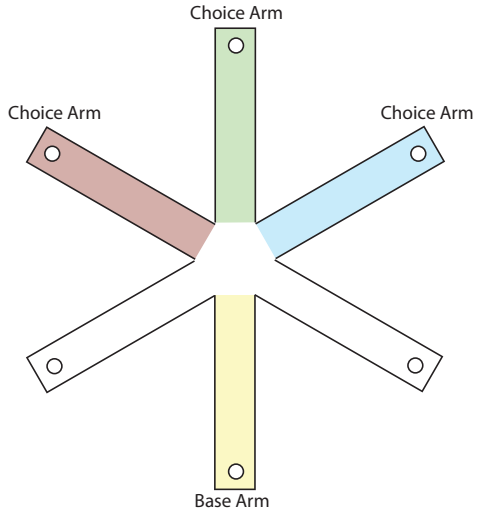

Fig. 1. *n*-armed Bandit Task maze. The task was conducted using a six-arm radial maze. Three arms were designated to be choice arms, while one arm served as the return to base arm. The remaining two arms were not cued and did not deliver any reward.

(Hamilton 75 RN, Reno, NV) was lowered to the appropriate stereotaxic coordinates. Injections were made at two sites in each hemisphere, both located within the PL region using the following coordinates: site 1: anterior–posterior (AP), +3.5; medial-lateral (ML), ±0.75; dorsoventral (DV) relative to dura, −3.3; site 2: AP, +2.5; ML, ±0.75; DV, −3.2. Injections of 0.4 μl of NMDA (15 mg/ml, Sigma-Aldrich, Oakville, ON) were controlled manually at a rate of 0.1 μl every 2 min with 4 min for absorption before the needle was retracted.

After removing the injector, the craniotomy was filled with gel foam and a topical antibiotic was applied after suturing to prevent infection. All rats were treated with Metacam (1 mg/kg; concentration: 5 mg/ml, Boehringer Ingelheim, ON) for three days post surgery at 24 h intervals. Animals remained in their home cages for 9 days following surgery to allow the animals to recover. During this time, food was available ad libitum for 4 days, after which, the rats were food restricted to 85% of their free-feeding weight. Water was available ad libitum for the entire duration of recovery.

2.3. Behavioural apparatus

Behavioural testing took place on a six arm radial maze (see Fig. 1), built in-house. The dimensions of the maze are as follows: maze diameter = 1.46 m, arm length = 0.61 m, arm width = 0.15 m. The entire maze was elevated 0.64 m above the floor. Three of the six arms were assigned to be choice arms and one arm was designated as the base arm. The two remaining arms were unlit and unrewarded. At end of each arm was a port capable of cuing the animal with a flashing light, registering a nose-poke, and delivering a high-calorie liquid reward (chocolate Ensure®, Abbott Laboratories, Ltd., Saint-Laurent, QC). The food well in each port was attached via silicon tubing (inner diameter 1.98 mm, outer diameter 3.18 mm, VWR International, Mississauga, ON) to a 50 ml syringe which was mounted to the back wall of each arm 65 cm above the track floor. The syringes contained a reservoir of liquid Ensure® which gravity fed through a solenoid pinch valve (Asco Scientific, model SCH284B004, Florham Park, NJ) and into the food well. The valve remained closed until the rat nose-poked an available port at which time the valve would open for a set amount of time (i.e. 300 ms, 750 ms, or 1200 ms, corresponding to approximately 0.038 ml, 0.139 ml, and 0.239 ml of Ensure®, respectively). Every port was fitted with a white LED to cue port availability and a horizontal infrared emitter-detector pair (Honeywell, models SD5610-001 & SE5470-004, Golden Valley, MN) that was used to sense rodent nose pokes into the food well. The experiment was controlled and data acquired via a National Instruments digital input/output board (PCIe-7841R, Toronto, ON) using custom-written Labview (version 10.1, National Instruments, Austin, TX) software on a standard Windows-based computer in an adjacent room. Video was recorded using a ceiling mounted video camera placed above the maze and video tracking was performed using Cheetah software (Neuralynx, Inc., Bozeman, MT) on a computer located in an adjacent control room.

2.4. Behavioural testing

2.4.1. Prior testing

After surgical recovery and prior to testing on the *n*-armed Bandit Task, all animals participated in a suite of tests investigating the role of mPFC in behavioural shifting using digging pots and odour cues, loosely following the methods of Birrell and Brown [19]. Briefly, this study required rats to detect a food reward (i.e., round shaped toasted oat cereal) that was hidden within one of two scented bowls covered either by corncob bedding material or silica sand. The location of the food reward was cued either by the scent in the bowl (e.g., coffee vs. blueberry) or by the digging media (e.g., corn cob vs. sand). The experiments involved a series of simple discriminations, reversals, and extradimensional shifts (shifting from using odour cues to digging media to identify the rewarded bowl). Each rat participated in these tests for roughly 30 min each day for 10 days before moving onto the *n*-armed Bandit Task. After this training, rats were given 2–6 days of recovery before starting the *n*-armed Bandit Task pre-training. Due to the completely different nature of the two tasks, we expect the influence of this prior testing on our results to be minimal.

2.4.2. Testing time and food deprivation

All habituation, pre-training and testing sessions took place between 11:00 AM. and 6:00 P.M. four to six days per week. Water was available *ad libitum*. Animals were food restricted to ~85% of their free-feeding weight and were maintained at this weight by daily supplements of standard rat chow, as needed, provided at the end of behavioural testing.

2.4.3. Habituation and pre-training

Before habituation began, each animal was given a small amount of chocolate Ensure® one to two days before pre-training began in order to create familiarity with the novel food. Animals were habituated to the radial maze one at a time in a single pre-training session (18–59 days post surgery [mean = 30.7 days]), during which all six ports were illuminated. Each port delivered 750 ms of food reward when nose-poked. Animals were then trained to make a nose-poke response into a single illuminated port. The spatial location of the stimulus light varied randomly between trials across 6 arms. Each session contained 150 trials and lasted approximately 20 min. Rats were moved onto the training stage either after they had completed 150 trials within a 20 min pre-training session or after completing ten pre-training sessions. Twenty rats achieved 150 trials within 20 min and three rats (1 lesion; 2 sham) were moved onto the training phase after ten sessions.

2.4.4. Training

Animals were trained over 21, twenty min sessions to flexibly seek out the port that offered the most reward. Animals were placed in the centre of the maze at the beginning of each session facing away from the base arm. Each trial required the animal to first nose poke the port at the end of the base arm, collect the 300 ms reward there, then turn around and travel to the middle

of the maze (i.e., the decision zone). At this point the rat could freely choose one of the three choice arms from which to collect reward. Each port offered a different amount of food reinforcement (300 ms, 750 ms, or 1200 ms) and time-out punishment (20 s, 10 s, or 0 s). The main purpose of the time-out periods was to enhance the discriminability of the reward amounts; during this time, rats were left unrestricted on the maze. The high reward arm (HRA) always offered a large amount (1200 ms) of reinforcement with no (0s) time-out, the medium reward arm (MRA) offered a moderate amount (750 ms) of reinforcement and a moderate (10 s) time-out, and the low reward arm (LRA) offered a small amount (300 ms) of reinforcement and a large (20 s) time out. Note that we use the term "reward" throughout this paper to indicate the overall value of each response option (i.e., the net effect of reinforcement and punishment). Through trial-and-error each rat was able to determine which of the three choice ports offered the largest reward. Once the animal had chosen the HRA ten times in a row, the experimenter would initiate a new trial block in which the locations of the HRA, MRA, and LRA were shuffled. New HRA, MRA, and LRA locations were chosen pseudo-randomly in that every arm changed food/punishment amount (i.e., one arm could not continue to offer the same amount of reward and punishment for multiple trial blocks). At this point, the experimenter helped train the rats manually to explore the different options. This involved the experimenter coming into the testing room and leading the rat to the HRA with her hand ~4–5 times and exiting the room. Care was taken to ensure that all rats received similar amounts of intervention and the majority of the interventions took place during the first half of training and were tapered off in the latter half. No interventions were allowed during the testing phase. After the 21 training sessions were completed, animals were moved onto the testing phase of the experiment. We deliberately included large numbers of training sessions to ensure that all animals had reached peak individual performance levels before moving onto the testing phase. This ensures that possible differences in learning due to unequal trial numbers between groups would have a minimal impact on performance levels during the testing phase.

2.4.5. *The* n*-armed bandit task*

The design of the *n*-armed Bandit Task was similar to the training phase described above in that rats attempted to obtain the largest amount of food reward possible by persisting at the HRA for a number of trials and then flexibly adapting to a change in HRA location when that arm became devalued after a switch. The main differences in this stage were that switches were no longer tailored to rodent performance and experimenter interventions were not permitted. Sessions were set up in such a way that switches occurred automatically after 35 (±5) trials. The timing of each switch was varied so that the rats would not be able to predict exactly when a switch was going to occur. Each rat completed 16 days of testing and each session was 20 min in length. Rats completed as many trials as possible in each session within the allotted 20 min time limit.

2.5. Histology

Following completion of behavioural testing, animals were sacrificed via an overdose of 100 mg/kg sodium pentobarbital, then transcardially perfused with 1x phosphate-buffered saline (PBS) and 4% paraformaldehyde (PFA). The brains were removed and post-fixed in 4% PFA for 24 h before being stored in a 30% sucrose solution with sodium azide. Brains were sliced into 40 μm sections using a cryostat and Nissl stained with 0.5% cresyl violet.

Sections were analysed using a microscope capable of imaging entire slides at 40× (Nanozoomer, Hamamatsu Photonics, Honshu, Japan). The extent of the lesions were manually mapped onto standardised sections of the rat brain [46].

2.6. Data analysis

2.6.1. Reinforcement learning model fitting

Each rat's behaviour was fit with a reinforcement learning model, which allowed us to quantify its choice behaviour [43,44]. In this model, the agent (rat) keeps a running estimate of the expected reward available on each arm, denoted $Q(c)_t$ for choice arm c on trial t. After each choice, the expected reward for the chosen arm on the next trial, $t+1$, is updated based on the difference between the actual reward, r_t, and expected reward:

$$Q(c)_{t+1} = Q(c)_t + \alpha\,[r_t - Q(c)_t] \quad (1)$$

The parameter α describes the learning rate (i.e., the weighting placed on the reward prediction error when adjusting expected value), which varies between 0 and 1. Low values indicate that the animal is adjusting its reward estimates very slowly, requiring integration of reward feedback over many trials before internal estimates will match the true values. High values indicate that the rat is updating rapidly based on the previous reward.

In order to translate estimated reward values into an actual choice, we used the softmax decision rule [43]:

$$P(c) = \frac{\exp\left(\beta Q(c)\right)}{\sum_c \exp\left(\beta Q(c)\right)} \quad (2)$$

here P(c) is the probability of choosing arm c, and the summation on the bottom is conducted over all values of c (in our case 1, 2 and 3). The parameter β, often referred to as the inverse temperature, describes how much the present choice will be weighted towards the highest value. If β is zero, the probability of each arm will be equal, independent of the expected values (i.e., choice will be completely random). As β approaches positive infinity, the softmax rule places a very high selection probability on the choice with the highest expected value and a very low probability on all other options (i.e., it becomes "winner-takes-all").

For all choices from each rat, the negative log likelihood was computed as the sum of the log of the probabilities, P(c), of all choices. A non-linear optimization routine (fminsearch in Matlab, the Mathworks, Natick, MA) was then used to minimize this quantity, yielding the optimal values of α and β for that animal [44].

2.6.2. Place and reward entropy

To measure the selectivity of each animal for either location or reward amount, we used an entropy measure. Within blocks of trials, we computed the proportion of times that the rat would choose a given maze arm location, p_a, independent of the reward available on that arm. The place entropy was then computed using:

$$\text{place entropy} = -1\sum_a p_a \log_2 p_a \quad (3)$$

where a varies from 1 (left arm) to 3 (right arm). Place entropy was computed for the entire course of testing for each animal, using a sliding window of 210 trials, advanced in steps of 10 trials. We chose blocks of 210 trials because this corresponds to, on average, 6 reward-value switches (average trials between switches is 35). The average entropy measure presented below was taken across all entropy measures computed for a given animal.

Reward entropy computations were similar except that the proportions used in the formula were based on the number of times the animal ran to an arm with a particular reward amount, independent of where that reward was available. Unlike with place entropy, we excluded the first 20 trials after each switch from this measure because the learning curves showed continued improvement

in reward discrimination during this period. All other details were identical.

2.6.3. Quantifying motor stereotypy with path entropy

To quantify the degree of motor stereotypy exhibited by our rats, we measured the lateral deviations in a rat's path as it crossed a specific location on the maze and computed the entropy of these deviations. Lateral deviations were manually encoded using video recordings from the first and last days of testing with the aid of an open-source video markup software package (Kinovea, version 0.8.15, www.kinovea.org). Specifically, for all outbound trajectories from the base port to one of the choice arms, data points indicating the lateral position of the rat's nose were marked along three lines orthogonal to the direction of travel. These three lines were evenly spaced along the base arm with the first roughly 3 cm from the base port and the last roughly 3 cm before the central hub of the maze. The position data were then split into three groups according to the arm that the rat chose on that trial. Within each choice, the positions were binned into 8 evenly spaced bins and expressed as proportions of all trials for that choice arm. The entropy of this distribution was then computed as:

$$\text{path entropy} = -1\sum_{a} p_b \log_2 p_b \quad (4)$$

where p_b is the proportion of trials in a given bin and the sum is taken over all 8 bins. The final path entropy measure for a given rat on a given session was the mean of the entropy measures for each choice arm, weighted by the number of times the rat actually choose the specified arm.

2.6.4. Correlating performance with anatomical location of damage

To determine whether damage to specific brain regions corresponded to particular functional deficits, we analysed the correlation between damage and both place and reward entropy across the extent of the mPFC. Lesions were first manual mapped onto normalized coordinates given by the Paxinos and Watson atlas [46], as we did previously when characterizing the extent of the lesions. Next, the lesion sites were scaled by rendering the images at 43 pixels/mm. Then each rat's left hemisphere lesion was reflected onto the right hemisphere under the assumption that any lesion effects would be bilaterally symmetric. Thus, there were two unilateral lesion maps for each animal in the final data set. For each pixel in this data set, we correlated the lesion status (1 or 0) with the place or reward entropy value computed for each animal.

3. Results

3.1. Lesions were centered on the PL region of mPFC

Superimposed images illustrating the extent of the mPFC lesions are provided in Fig. 2. Excitotoxic lesions were targeted at the prelimbic (PL) region, and that region was nearly completely lesioned in the majority of animals. The damage also extended into adjacent regions, including the infralimbic cortex (IL), medial orbital cortex (MO), and the anterior portions of both the secondary motor cortex (M2) and cingulate cortex (CG1) in several rats. Slight damage to the ventral orbital cortex and posterior cingulate cortex (CG2) was observed in a minority of animals.

3.2. Rats with mPFC lesions are less adept at adjusting to changing reward contingencies

To assess the ability of lesion and control animals to adjust to changes in reward value among the three arms, we calculated the amount of reward obtained on each trial after a switch, averaged over all switches for that subject. As expected, all animals achieved much lower reward on the first trial after a switch followed by a gradual adjustment over the next 20–30 trials (Fig. 3C). We analysed the rate of adjustment by fitting this curve with an exponential function by taking the natural log of the x-axis (trials after switch) and performing a linear regression over the first 30 post-switch trials. Slopes obtained via this curve-fitting method were significantly lower for the lesioned rats than the control animals ($t(20, 22) = 3.077$, $p = 0.006$, shown in Fig. 3D), showing that lesioned animals were slower to adjust to changing reward contingencies. Further, lesioned animals reached lower asymptotic levels of reward, as shown by a two factor repeated measures ANOVA comparing the amount of reward received during trials 20–30 by lesion and control animals. The analysis revealed a significant within subjects effect of trial ($F(1, 10) = 5.446$, $p < 0.001$) and a significant between subjects effect of lesion ($F(1, 20) = 10.293$, $p = 0.004$), the interaction of trial x lesion was not significant ($F(1,10) = 1.584$, $p = 0.113$). Thus, lesion and control animals differed in overall level of reward achieved. Further, both groups were still showing minor performance improvements 20–30 trials after the switch. Overall, the results show that mPFC lesions cause both transient impairments in adjusting to a switch in reward contingencies and a persistent pattern of suboptimal choices long after the switch.

Because lesion animals choose lower reward more frequently, they might therefore have received less testing trials, creating a potential confound in our analysis. However, we compared the total number of trials and, although lesion animals had fewer trials overall, the difference between groups was not significant (average trials per session for control ± standard error of the mean, SEM: 80.4 ± 4.0; lesion: 72.7 ± 3.4; *t*-test: $t(20) = 1.4$, $p = 0.18$). To investigate this issue further, we also compared performance on the first half of all testing trials to that on the second half of all testing trials. A repeated measures *t*-test between first and second halves showed no significant differences in the slope of the post-switch reward curve (i.e., Fig. 3C) for either controls (means ± SEM: first half = 9.1 ± 0.7, second half = 8.9 ± 0.7; $t(12) = .19$, $p = 0.85$) or lesioned animals (means ± SEM: first half = 5.8±0.8, second half = 5.3 ± 1.5; $t(8) = 0.50$, $p = 0.63$). Similarly, we found no significant differences in the amount of reward obtained during the asymptotic phase (i.e., Fig. 3C, trials 20–30) for either controls (means ± SEM, expressed as percentage of maximum reward: first half = 80.4 ± 1.4%; second half = 82.6 ± 1.5%; $t(12) = -1.89$, $p = 0.08$) or lesioned animals (means ± SEM: first half = 73.0 ± 2.5%, second half = 73.8 ± 2.9%; $t(8) = -0.39$, $p = 0.71$). These results show that rats' performance were stable across 16 days of testing, suggesting that they had reached asymptotic performance during the training phase. Therefore, any extra trials that control animals received should not have affected the overall pattern of results.

3.3. Rats with mPFC lesions adjust reward estimates more slowly but show no difference in exploration

We further investigated the effects of mPFC lesions on reward-guided decision-making by fitting each rat's performance with a reinforcement learning model. As described in the introduction, these models are based on the premise that the rat maintains a running estimate of the reward available at each zone and adjusts these values after each decision according to the reward prediction error. The rat then generates a choice based on these estimated values that is probabilistic and biased towards what it thinks is the choice with the highest value. Reinforcement learning models provide insights into parameters of decision-making that are not otherwise directly observable [44]. In the present model, explained in detail in the Material and Methods, the parameter α provided a measure of the rate of adjustment of internal value estimates based on reward feedback and is often referred to as the "learning rate". It ranges

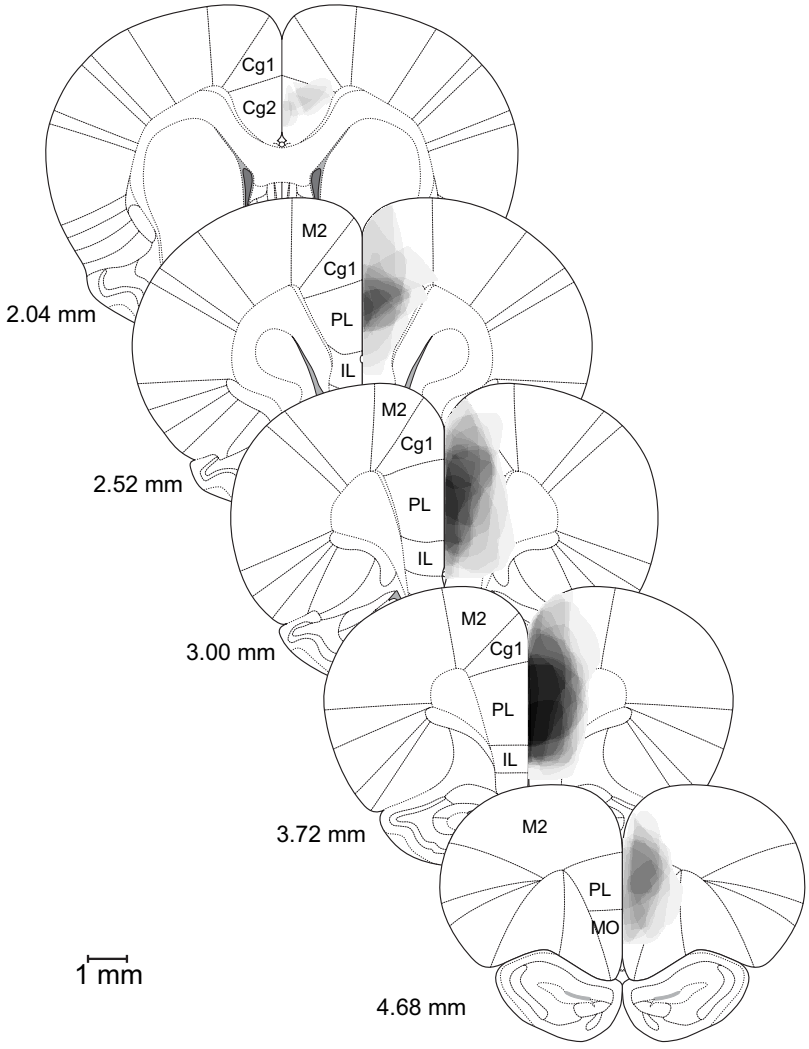

Fig. 2. Superimposed images of lesion extent mapped onto standardised sections of the rat brain for all lesion animals. Lesions were bilateral but have been projected onto left hemisphere for clarity. Darkest areas indicate the largest overlap of damage in lesion animals. Distance of sections from bregma is indicated in lower left. Adapted from Paxinos and Watson [46].

between 0 and 1, where high values mean the rat adjusts rapidly based on only the last 1–2 trials and lower values indicate a slower adjustment over a much wider range of trials. The other parameter, β, indexes the randomness of the choice and is sometimes referred to as the inverse temperature. 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. At the other extreme, when β is zero all choices are equally distributed, independent of expected value. Importantly, at intermediate values, a rat that explores more will exhibit more choices of suboptimal arms and will hence have a lower β value. Thus, β provides one way to quantify exploratory behaviour.

We fit each rat's choice performance over all trials during the testing period to the reinforcement learn model shown in Eqs. (1) and (2), yielding estimates of α and β for all animals. As shown in Fig. 4A, lesioned animals had lower α values than control animals ($t(17.56, 22) = 4.78$, $p < 0.001$; equal variances not assumed), suggesting that they are slower to adjust internal reward estimates based on feedback. This is also completely consistent with their slow adjustment to changing reward contingencies, as shown in Fig. 3C. To a large extent, these two analyses tap into the same underlying ability. Interestingly, though, we did not observe a significant difference in our measure of decision randomness, β (Fig. 4B, $t(9.087, 22) = -1.87$, $p = 0.094$; equal variances not assumed). It should be noted that the group difference in β trended towards significance; however, this appears to be driven largely by two animals with a strong tendency to perseverate on a single choice arm (discussed further below). The range of β estimates from the remaining animals completely overlaps with that of the controls. Hence, mPFC lesions impair internal updates of reward value but appear not to affect the randomness of choices and, by extension, have no effect on the tendency to explore.

As an alternate measure of exploratory behaviour, we counted the number of trials in which the animal (1) choose an arm with less than the maximum available reward and (2) chose an arm different than on the previous trial. For both control and lesioned animals combined, roughly 15 percent of trials fit this definition of an exploratory trial. Again, we found no differences between groups (control (mean ± SEM): 15.9 ± 0.1%; lesion: 14.3 ± 0.1%; $t(18) = 0.48$, $p = 0.64$). In sum, neither our model-fit exploration parameter, β, nor a simple classification of exploratory trials showed a significant difference between groups.

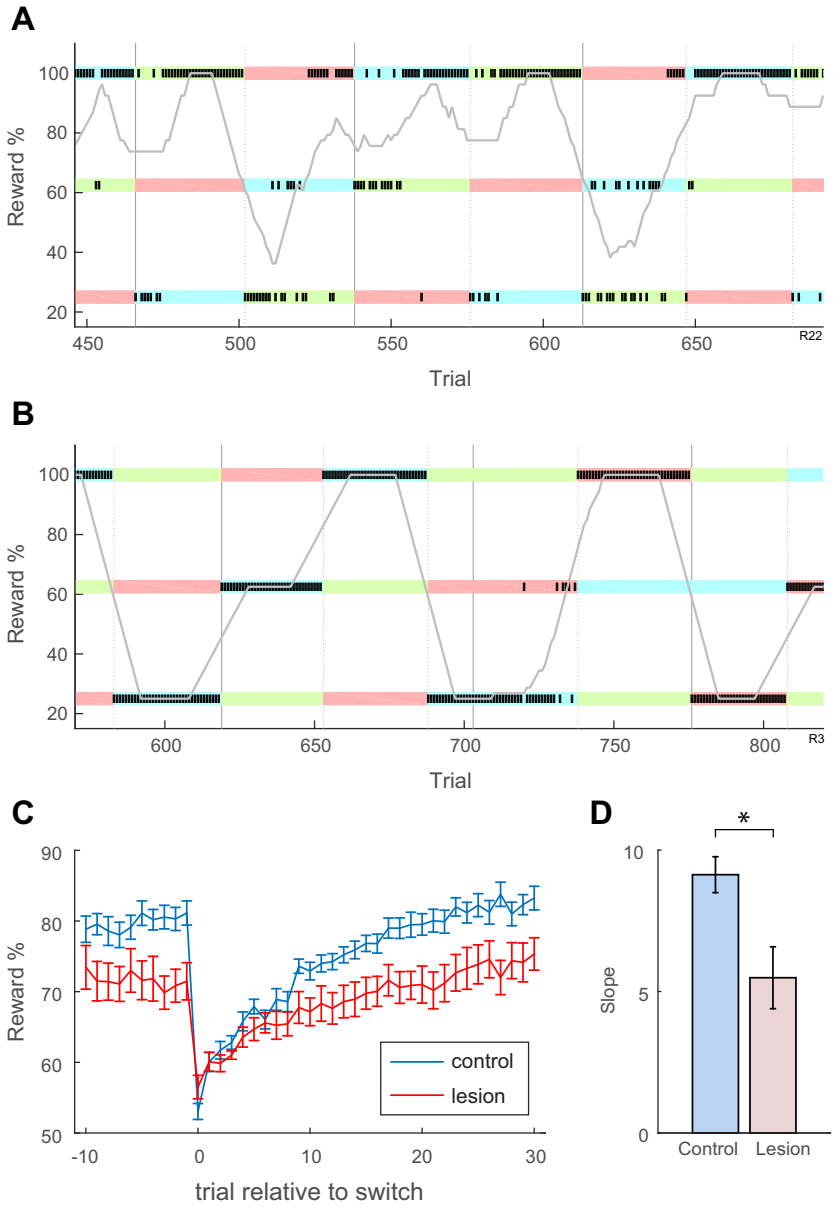

Fig. 3. Effects of mPFC damage on switch performance. Plots show (A) a typical control animal's performance and (B) the most impaired lesion animal's performance across 250 reward trials (approximately three sessions). Responses on each trial (small black tick marks) are plotted as a function of the reward received on that trial (expressed as a percentage of the maximum available reward). Note that the location (i.e. arm) at which a given reward amount was received switched every ~35 trials. Hence, the physical location of the animal's response is shown by the pastel colour band behind the tick marks (pink = left arm, green = middle arm, blue = right arm). Solid vertical gray lines indicate the end of one daily testing session and the beginning of a new one (i.e., all trials between two such lines were collected in a single session). Vertical dashed lines indicate the point where reward values were switched within a session. The fluctuating horizontally-oriented grey line represents a moving average of attained reward with a window size of 20 trials. While the control animal switches choices so as to optimise the value of obtained reward, the lesioned animal shows perseveration on first the right arm (blue) and then the left (pink), which are independent of value. (C) Average reward received by lesioned and control animals as a function of trial position relative to a switch of reward contingencies. Data are first averaged over all contingency switches over all sessions for a single rat before averaging across animals. (D) Slope of the exponential curve fit to both lesion and control data shown in part C. Plots C and D illustrate that rats with mPFC lesions adjust more slowly to changing reward contingencies. Error bars represent the standard error of the mean (SEM). Significant effects ($p < 0.05$) are denoted with an asterisk. (In the black and white version of print, dark gray = left arm, light gray = middle arm, and white = right arm).

3.4. *mPFC lesions shift decisions towards place and/or response*

As shown in Fig. 3B, some lesioned animals showed a strong tendency to perseverate on a single choice arm, independent of the reward available on that arm. While the illustrated data are an extreme example, they never-the-less lead us to explore the selectivity for place using a measure we call place entropy (using Eq. (3)). Place entropy quantifies the randomness of choices relative to place, ranging between 0, meaning the rat choose the same maze arm on every single trial, to a max of 1.58, meaning that the rat chose randomly with respect to place. Note that due to counterbalancing of reward location across maze arms, an animal driven exclusively by reward value should distribute their choices relatively randomly with respect to place and should hence have a high value of place entropy. For the purposes of comparison, we also computed the reward entropy, which is similar to place entropy except that it indexes the randomness of choices relative to reward value (see Eq. (3) and accompanying text). Note that in terms of task performance, these two entropy measures should be inversely related. A

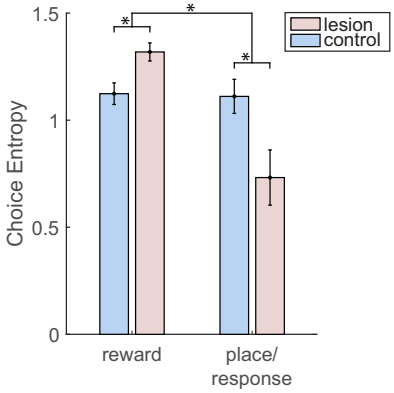

Fig. 5. Effect of mPFC lesions on reward and place/response entropy measures. Lesions of mPFC produced a significant increase in reward entropy and a significant decrease in place entropy, showing that choices in lesioned animals were more influenced by place or turn direction rather than reward value. Error bars represent the SEM. Significant effects ($p < 0.05$) are denoted with an asterisk.

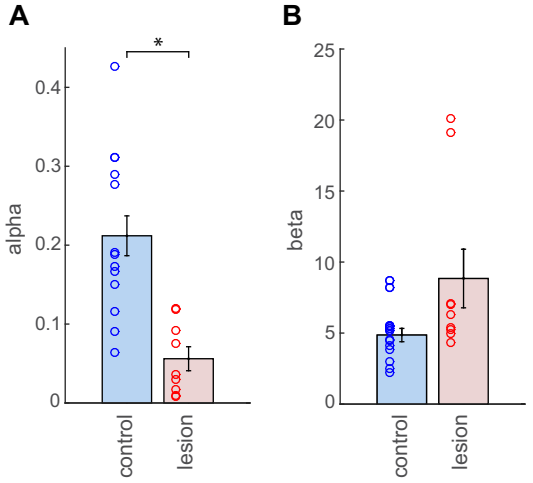

Fig. 4. Effect of mPFC lesions on (A) learning rate (α) and (B) inverse temperature (β). The presence of mPFC damage lead to lower α values but no significant difference in β values. Circles depict data from individual rats. Error bars represent the SEM. Significant effects ($p < 0.05$) are denoted with an asterisk.

rat making optimal choices would have a very low reward entropy and very high place entropy.

We computed place and reward entropy in blocks of 210 trails (covering 6 switches in reward location), using a sliding window over the entire history of *n*-armed Bandit choices for that animal. Average entropy values were then computed for each animal across all trials. As shown in Fig. 5, mPFC lesions caused changes in both reward and place entropy. Independent samples *t*-tests showed that mPFC lesions had a significant effect on both reward entropy ($t(20) = -2.28$, $p=0.034$) and place entropy ($t(20) = 2.65$, $p=0.015$). While control animals had almost the same reward and place entropy values, lesioned animals were shifted towards higher reward entropy and lower place entropy. These results show that lesioned animals were more random with respect to reward value and more likely to be biased towards particular reward arm locations.

Given that an animal responding optimally on the task would have reward entropy values near zero, the relatively high reward entropy values for the control group were puzzling. We know from Fig. 3C that both groups of animals obtained much more reward than expected by chance and so were reward-sensitive. Closer inspection of the choice behaviour of individual rats revealed that many animals, even in the control group, were strongly biased away from one arm. When this behaviour was evident, the animal would still choose the higher of the remaining two arms, but would not switch to the avoided arm. The fact that place entropy was considerably below the maximum value in both lesion and control animals shows that place is a strong determinant of our rats' choice behaviour in this task.

Importantly, the literature suggests that place-based learning and response-based learning (i.e., turn direction) are driven by separate neural systems [47]. Because rats in our task always made the same turn direction to each choice arm, we cannot distinguish between these two strategies and so our place bias may mean either a strong preference for place or a strong preference to make a particular turn direction. Accordingly, we have used the term place/response entropy in Fig. 5.

3.5. mPFC lesions do not affect motor stereotypy

Given the strong tendency of rats with mPFC lesions to perseverate on place/response, we tested the hypothesis that lesions would also affect flexibility in motor execution. As habits develop, behaviour typically becomes more stereotyped [48,49]. Very little research has been conducted that directly compares stereotyped choices and motor behaviour. However, Schmitzer-Torbert and Redish [48] investigated the development of path stereotypy using a multiple-T task and found that after rats learned the correct response sequence, stereotyped motor sequences routinely emerged, but only after many passes through the maze. The parallel emergence of stereotyped choices and stereotyped motor patterns raises the possibility that these two forms of stereotypy might depend on a common neural substrate. Given that our data show that mPFC lesions lead to more stereotyped choices, we wondered whether they might also lead to concurrent increase in motor stereotypy, measured as reduced variance in their paths. Accordingly, the horizontal deviation of each rat's path was measured at three positions on the base arm as the rat made an outbound journey towards one of the three reward arms. An entropy measure was then used to quantify the randomness of the paths, with lower entropy values meaning more stereotyped behaviour (see Eq. (4)). These entropy measures were computed for both the first and last days of testing, allowing us to examine the development of stereotypy as a result of training. A two-factor mixed method ANOVA was used to assess the impact of lesion and training (first day versus last) on path entropy. At the first and second positions (the two positions closest to the base arm reward port), we found no significant differences due to lesion or training. At the third position (the one furthest from the reward port and closest to the choice), there was a significant main effect of training day (Fig. 6, $F(1, 20) = 17.179$, $p = 0.001$), showing that behaviour became more stereotyped due to training. These results validate our measure as they show the expected pattern of increased stereotypy with increased training. Surprisingly, we failed to find any difference between groups, $F(1, 20) = 2.406$, $p = 0.137$, or interaction (treatment $\times$ session), $F(1, 20) = 1.641$, $p = 0.215$. Furthermore, the Pearson correlation coefficient was computed to determine whether a relationship was present between path entropy and place entropy measures. There was no significant relationship between these two measures, $r^2(20) = 0.002$, $p = 0.85$. In sum, these findings suggest that although mPFC damage increased stereotyped decision-making, this effect was not reflected in rats' motor behaviour.

3.6. Deficits depend on the brain areas damaged

Our analyses of reward and place entropy showed a wide range of disabilities within the lesioned group, ranging from almost

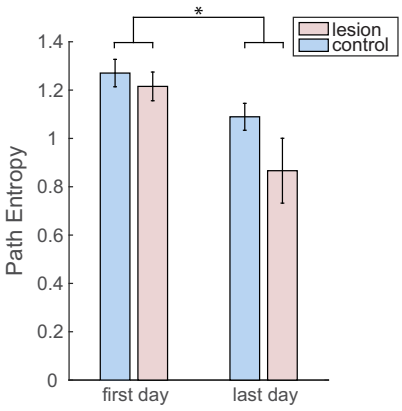

Fig. 6. Effect of mPFC lesions and training on motor stereotypy. A comparison of path entropy measures obtained for lesion and control animals on the first and last day of testing for the horizontal position closest to the centre of the maze. Both groups of animals exhibit an increase in stereotyped behaviour during the last day of testing compared to the first day. Error bars represent the SEM. Significant effects ($p < 0.05$) are denoted with an asterisk.

normal to severely impaired. Given that we also had some variability in the extent and spatial patterns of our lesions (see Fig. 2), we considered whether damage to certain brain areas was predictive of particular patterns of impairment. To address this question, we looked for correlations between damage and performance (i.e., place and reward entropy) for every brain location within our group of lesioned animals. For this analysis, we included all animals in our lesion group, irrespective of lesion size, for a total of 12 animals. Pearson r^2 values for place and reward entropy are shown in Fig. 7. Note that with 12 animals, r^2 values above +/− 0.29 are significant ($p < 0.01$). As shown in Fig. 7A, animals with damage extending into the posterior region of the dorsal anterior cingulate cortex were more likely to repeatedly choose one place and/or response. In contrast, rats with damage in the anterior prelimbic region were most likely to show random choices relative to reward value (Fig. 7B). There was a tendency of animals with damage extending into a focal region in the medial infralimbic region to actually be better on both reward and place entropy measures, but this may simply reflect the fact that animals with lesions in this deep region were less likely to have lesions in more dorsal regions strongly associated with impairment. It is important to note that these results are based on a very small number of animals with largely overlapping lesions. Thus, we consider these results suggestive, but not conclusive. We include it because the pattern of results leads to some intriguing ideas concerning the role of the ACC in behavioural flexibility that we believe merit further investigation.

4. Discussion

Our data indicates that, in rats, damage to the PL region of the mPFC alters foraging-type decision-making in several different ways. Firstly, animals with mPFC lesions have difficulty with adjusting behaviour so as to optimise reward. This effect was illustrated by the lesioned rats' decreased ability to shift to the high reward arm when reward values shifted. This behaviour also manifested as lower learning rates in our reinforcement learning model fitting. Asymptotic performance levels were also lower in this group, as shown by their overall lower levels of obtained reward and the increased randomness of their choice relative to reward (i.e., higher reward entropy). Secondly, and contrary to our expectations, mPFC damage did not have a major impact on the ability of rats to engage in exploration, a behaviour we quantified via the inverse temperature measure (β) obtained from our reinforcement learn-

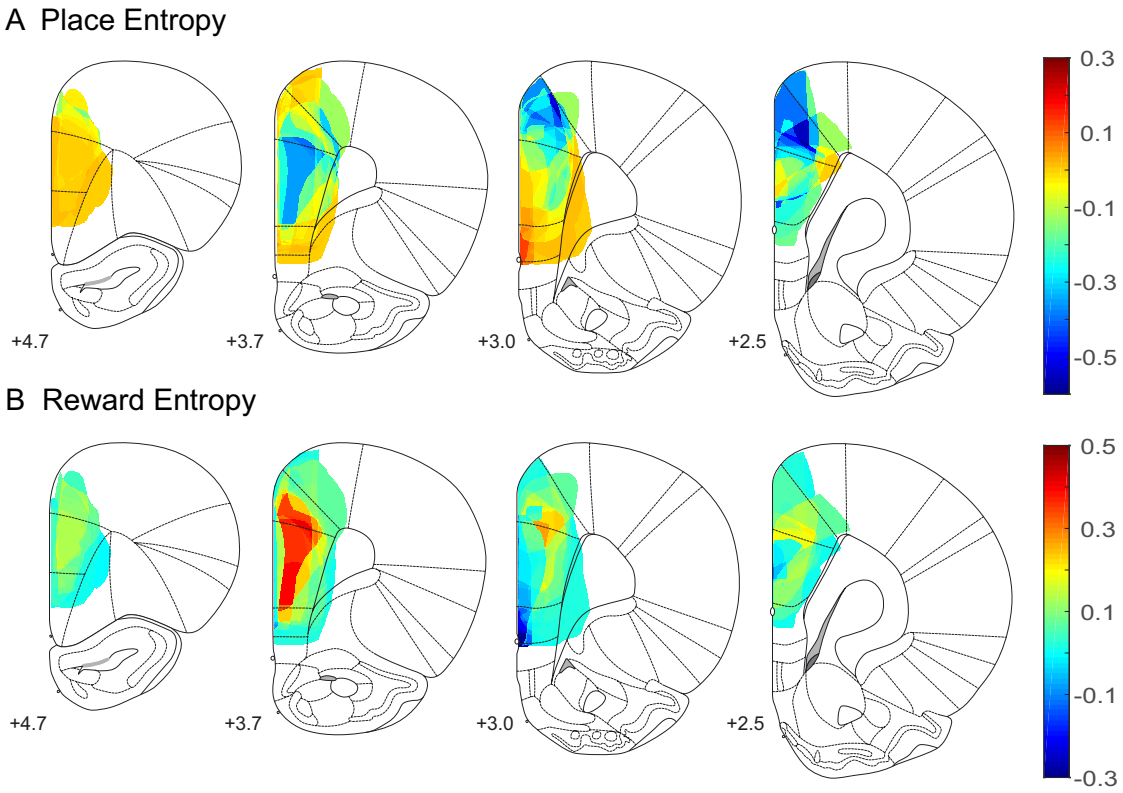

Fig. 7. Region specific effects mPFC damage on place and reward entropy. (A) Place entropy is inversely correlated with damage to the posterior region of the dorsal ACC, meaning that animals with damage to this region paid more attention to place. (B) Reward entropy is positively correlated with damage to anterior PL, meaning that animals with damage to this region were more likely to respond randomly with respect to reward value. r^2 values above +/− 0.29 are significant ($p < 0.01$).

ing model fitting. That is, lesioned and control animals exhibited comparable tendencies to choose arms randomly rather than exploiting the arm with the highest known reward value. Lastly, lesioned animals displayed perseverative biases in their choice behaviour, but not in their motor behaviour when compared to control animals. Specifically, spatial locations on the maze appeared to factor more heavily into the decision-making processes of rats with mPFC injury and they thus exhibited less choice randomness in relation to place than intact animals (i.e., lower place entropy). Correlating both reward and place entropy values with regional damage revealed a possible dissociation between reward-tracking – mediated by the anterior PL region and response flexibility – mediated by dorsal mPFC (i.e., primarily the CG1 region). However, contrary to our expectation, animals with strong perseveration on place were no more likely to show stereotyped paths than other animals, suggesting that choice stereotypy and motor stereotypy may be mediated by separate neural systems.

As previously mentioned, one of our primary motivations was to examine the role of mPFC in exploratory decisions (i.e., choices where the rat forgoes an expected high-value reward for the possibility of discovering an even higher reward elsewhere). We measured exploratory behaviour primarily by fitting each rat's choices with a reinforcement learning model, and using the β parameter from our softmax decision rule. Beta describes how the deciding agent (in this case, a rat) might transform its internal estimates of the reward value available from each arm into the probability of choosing a given arm. Low β values indicated an animal that frequently made choices other than the highest value option, suggesting more exploration [43]. There was no statistically significant difference in β values between our control animals and those with mPFC lesions centered on the PL region. Thus, while PL damage did disrupt the ability of rats to flexibly adapt to changing task demands, it did not cause significant impairment in the ability of rats to engage in exploration, at least by this measure. One caveat with this measure of exploration is that we cannot distinguish between intentional choices to explore and pure randomness. Exploration usually means forgoing a known high reward in order to gain information about other possibilities. However, in simple reinforcement learning models like the one used here, exploration is implemented by introducing random choices. Strictly speaking, we cannot say whether PL lesions lead to changes in information seeking, only that lesioned and control animals were equally random in their choices. Studies using fMRI in humans have identified both the frontopolar cortex and dorsal anterior cingulate as playing a role in the explore-exploit decision-making [40,42]. Our data suggest that the rat PL is not homologous to either of these regions. However, it is worth noting that two of our lesioned animals showed very high β values, indicating serious impairment in exploration. These animals had damage that extended dorsally into ACC regions, possibly indicating a role for rat ACC in explore-exploit decisions.

While we failed to show an influence on explore-exploit decisions, our data did show that rats with lesions centred on PL were impaired in decisions between varying amounts of reward. On the surface, our task would seem to be closely related to reversal learning (i.e., rats must respond to one reward option while inhibiting the response to a previously rewarded option). Therefore, the observed switching impairment due to PL damage appears to conflict with several studies which show that rats with PL damage are not impaired on reversal learning [50,51], or are only transiently impaired [52,53]. Instead, reversals seem to depend on the orbitofrontal region [34,51,54–58]. However, our task has two features which differentiate it from reversal learning tasks. First, because we use graded reward values, at least some reward was available on every arm on every trial. Reversal tasks, in contrast, typically involve a black and white switch (i.e., a complete loss of reward on the original choice). Second, our task involves three options while choices in reversal tasks are typically binary. It would thus appear that rats with PL lesions are impaired when the choice involves either graded reward, multiple options, or a combination of the two. Given that PL lesions typically do not impair free foraging on a multi-arm radial maze [9,14], it seems likely that the functional impairment involves discriminating among different values of reward.

Our findings of impaired value-based decisions after mPFC lesion fit well with several lines of evidence implicating mPFC in "goal-directed" actions. It has been suggested that the dorsomedial striatum mediates early phases of learning, where behaviour is concerned with achieving goals and therefore highly sensitive to outcome values [59]. The dorsolateral striatum, in contrast, has been suggested to mediate later stages of learning during which behaviours become stereotypical and habit-like and therefore insensitivity to outcome value [59,60]. Anatomical tracing data show that the PL region projects to the medial portion of striatum, thereby implicating it in goal-based actions [61,62]. In support of this supposition, several studies have shown that rats with damage centred in the PL region are impaired in decisions in which two actions lead to two types of reward, one of which has been devalued [31–33]. These findings are also consistent with the numerous rodent electrophysiological studies showing coding of the value of both expected and received reward by neurons in the PL region [26,63–66]. Thus, our findings, in conjunction with the existing literature, point to an important role for mPFC in either discriminating among different amounts of reward or attributing reward value to specific response options.

Our use of a three-choice task allows partial dissociation of response flexibility and reward tracking. Typically, value-based decision-making in rodents is studied with binary tasks [e.g. 30,54,67,68]. In a binary choice, a decision to switch from low- to high-value response ports also means switching from one response port to the other. On the other hand, in our task animals switching away from a low-value option could choose between two alternative locations. Our data show that, surprisingly, even normal rats often opted for a medium as opposed to high value reward in order to avoid one arm. Using our three-choice task, we were surprised to observe how strongly place determined choices in what should be, in theory, an entirely value-driven task. An optimally performing animal would have a place entropy of 1.58, indicating uniform distribution of choices across all three choice arms, and a reward entropy of 0, indicating responses exclusively on the high-reward arm. Instead, we found that control animals had roughly equal reward and place entropy values. We also observed this propensity during pre-training: rats naturally avoided certain arms and would even become locked into preferring one choice arm for extended periods of time (i.e., over multiple sessions). In fact, during pre-training (but not testing) we even tried to break rats of this tendency by actively guiding them into avoided arms. Despite our best efforts, place was still a major factor in most rat's choices. Lesions of mPFC made rats even more biased to respond based on place. For lesioned animals, place entropy was lower than reward entropy, showing that place was actually a stronger determinant of a rat's choice than reward value. Others have reported difficulty switching away from a place-based response after mPFC lesions, though these studies contrasted place responding with response-based strategies [69,70]. In some of these previous studies it has also been suggested that the underlying deficit is not so much the inability to inhibit place responding but rather the inability to flexibly *change* response strategies [50,71]. The fact that our animals were worse after a switch of the high-reward arm position is certainly consistent with this hypothesis. However, given the strength of the place responding, it also remains possible that mPFC lesions

cause place bias because they lessen the discriminability of reward amounts or weakened action-reward associations.

Our post-hoc analysis of the correlations between regional damage and functional impairments (Fig. 7A) suggested a special role for dorsal mPFC (i.e., CG1) in avoiding the magnetic attraction of place. Given the small number of animals with lesions extending into this region, these data are far from conclusive. However, others have reported spatial perseveration after ACC lesions. De Bruin et al. [69,70] found that rats with ACC lesions had difficulty learning a response-based strategy in a water maze (e.g., platform location is on the left), instead repeatedly swimming to the last experienced platform position in room-centered coordinates. Similarly, Seamans et al. [14] found that temporary inactivation of ACC, but not PL, lead to increased re-entries into previously rewarded arms of an 8-arm radial maze. Finally, Walton et al. reported mild perseveration on a delayed match to sample task in a T-maze [22]. More commonly, however, the ACC is implicated in effort-reward decision-making [21,23,72] and fear learning [73–75]. The finding that dorsal ACC damage was more detrimental to performance on our task than PL damage was particularly unexpected. The PL projects to the medial dorsolateral striatum, a region implicated in goal-directed actions [61,62]. The ACC, in contrast, projects to a section of dorsal striatum midway between the medial and lateral extremes [61,62]. The lateral striatum is strongly tied to habitual responding. In fact, damage to this area causes behaviours to become more sensitive to action-outcome contingencies [59]. Given this background, we expected that ACC lesions would, if anything, shift behaviours to be more goal-directed. Instead, we found that animals with damage extending into ACC were heavily influenced by place and largely insensitive to reward. Recent evidence suggests that the mapping of medial and lateral striatum into goals and habits, respectively, may be oversimplified. For example, recent data from a competitive choice task shows that lesions of lateral dorsal striatum decrease behavioural responses to losses more than lesions of medial dorsal striatum [76]. Further, electrophysiological data have shown that, among all frontal regions, the one with the strongest encoding of the value of an upcoming choice is actually M2 (medial agranular cortex), which projects to a striatal zone very close to the lateral striatum [61,62,77]. Hence, the poor choices of rats with lesions extending into ACC may reflect a breakdown in the ability to represent the value of upcoming actions, resulting in a regression to a place-based strategy that is largely insensitive to outcome values.

5. Conclusions

In sum, our results suggest that damage to the mPFC causes impairments in animals' ability to learn from feedback and use that information to guide future behaviour. In this regard, the findings are generally consistent with a role for the PL region in goal-directed action. We did not find any evidence that the rat PL region plays a role in explore-exploit decisions, although it may be that damage to more dorsal regions does. Using three choices, we were further able to show that place and/or response plays an important role in rat decisions on a radial maze, even when this bias leads rats to suboptimal reward collection. Further, damage to mPFC exacerbated this place bias. Finally, using a novel analysis technique, we were able to show a correlation between dorsal mPFC damage and place-based responding, suggesting an important role of the ACC and/or M2 regions in guiding actions based on action values. As this last finding is preliminary, we hope that further studies will expand on our mapping technique to better identify functional diversity within the frontal cortex.

Acknowledgements

This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC) and Alberta Innovates Health Solutions (D.R.E.). Additional support was provided via grants from the Alberta Gambling Research Institute (D.R.E. and R.J.W.). The authors also wish to thank Nathan House, Geoff Minors, Jenn VanOyen, Aleigha Arksey, Hilarie Swiftwolfe, Erica Nordin, Victoria Holec, and Dr. Matt Tata for their contributions to this work including maze construction, programming, data collection, and procedural advice.

References

[1] A. Bechara, A.R. Damasio, H. Damasio, S.W. Anderson, Insensitivity to future consequences following damage to human prefrontal cortex, Cognition 50 (1994) 7–15.
[2] G. Northoff, S. Grimm, H. Boeker, C. Schmidt, F. Bermpohl, A. Heinzel, et al., Affective judgment and beneficial decision making: ventromedial prefrontal activity correlates with performance in the Iowa Gambling Task, Hum. Brain Mapp. 27 (2006) 572–587.
[3] J. Reuter, T. Raedler, M. Rose, I. Hand, J. Glascher, C. Buchel, Pathological gambling is linked to reduced activation of the mesolimbic reward system, Nat. Neurosci. 8 (2005) 147–148.
[4] J.R. St Onge, S.B. Floresco, Prefrontal cortical contribution to risk-based decision making, Cereb. Cortex 20 (2010) 1816–1828.
[5] J.R. St. Onge, C.M. Stopper, D.S. Zahm, S.B. Floresco, Separate prefrontal-subcortical circuits mediate different components of risk-based decision making, J. Neurosci. 32 (2012) 2886–2899.
[6] M. Rivalan, E. Coutureau, A. Fitoussi, F. Dellu-Hagedorn, Inter-individual decision-making differences in the effects of cingulate, orbitofrontal, and prelimbic cortex lesions in a rat gambling task, Front. Behav. Neurosci. 5 (2011) 22.
[7] L. de Visser, A.M. Baars, J. van 't Klooster, R. van den Bos, Transient inactivation of the medial prefrontal cortex affects both anxiety and decision-making in male wistar rats, Front. Neurosci. 5 (2011) 102.
[8] L. de Visser, A.M. Baars, M. Lavrijsen, C.M. van der Weerd, R. van den Bos, Decision-making performance is related to levels of anxiety and differential recruitment of frontostriatal areas in male rats, Neuroscience 184 (2011) 97–106.
[9] B. Delatour, P. Gisquet-Verrier, Prelimbic cortex specific lesions disrupt delayed-variable response tasks in the rat, Behav. Neurosci. 110 (1996) 1282–1298.
[10] N.K. Horst, M. Laubach, The role of rat dorsomedial prefrontal cortex in spatial working memory, Neuroscience 164 (2009) 444–456.
[11] D. Durstewitz, N.M. Vittoz, S.B. Floresco, J.K. Seamans, Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning, Neuron 66 (2010) 438–448.
[12] E. Hartstra, J.F. Oldenburg, L. Van Leijenhorst, S.A. Rombouts, E.A. Crone, Brain regions involved in the learning and application of reward rules in a two-deck gambling task, Neuropsychologia 48 (2010) 1438–1446.
[13] E.L. Rich, M. Shapiro, Rat prefrontal cortical neurons selectively code strategy switches, J. Neurosci. 29 (2009) 7208–7219.
[14] J.K. Seamans, S.B. Floresco, A.G. Phillips, Functional differences between the prelimbic and anterior cingulate regions of the rat prefrontal cortex, Behav. Neurosci. 109 (1995) 1063–1073.
[15] K.A. Corcoran, G.J. Quirk, Activity in prelimbic cortex is necessary for the expression of learned, but not innate, fears, J. Neurosci. 27 (2007) 840–844.
[16] J.C. Churchwell, A.M. Morris, N.D. Musso, R.P. Kesner, Prefrontal and hippocampal contributions to encoding and retrieval of spatial memory, Neurobiol. Learn. Mem. 93 (2010) 415–421.
[17] C.M. Teixeira, S.R. Pomedli, H.R. Maei, N. Kee, P.W. Frankland, Involvement of the anterior cingulate cortex in the expression of remote spatial memory, J. Neurosci. 26 (2006) 7555–7564.
[18] D.R. Euston, A.J. Gruber, B.L. McNaughton, The role of medial prefrontal cortex in memory and decision making, Neuron 76 (2012) 1057–1070.
[19] J.M. Birrell, V.J. Brown, Medial frontal cortex mediates perceptual attentional set shifting in the rat, J. Neurosci. 20 (2000) 4320–4324.
[20] M.E. Ragozzino, The contribution of the medial prefrontal cortex, orbitofrontal cortex, and dorsomedial striatum to behavioral flexibility, Ann. N. Y. Acad. Sci. 1121 (2007) 355–375.
[21] V. Holec, H.L. Pirot, D.R. Euston, Not all effort is equal: the role of the anterior cingulate cortex in different forms of effort-reward decisions, Front. Behav. Neurosci. 8 (2014) 12.
[22] M.E. Walton, D.M. Bannerman, K. Alterescu, M.F. Rushworth, Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions, J. Neurosci. 23 (2003) 6475–6479.
[23] S.B. Floresco, S. Ghods-Sharifi, Amygdala-prefrontal cortical circuitry regulates effort-based decision making, Cereb. Cortex 17 (2007) 251–260.
[24] B.Y. Hayden, M.L. Platt, Neurons in anterior cingulate cortex multiplex information about reward and action, J. Neurosci. 30 (2010) 3339–3346.

[25] S.L. Cowen, G.A. Davis, D.A. Nitz, Anterior cingulate neurons in the rat map anticipated effort and reward to their associated action sequences, J. Neurophysiol. 107 (2012) 2393–2407.
[26] W.E. Pratt, S.J. Mizumori, Neurons in rat medial prefrontal cortex show anticipatory rate changes to predictable differential rewards in a spatial memory task, Behav. Brain Res. 123 (2001) 165–183.
[27] C. Amiez, J.P. Joseph, E. Procyk, Reward encoding in the monkey anterior cingulate cortex, Cereb. Cortex 16 (2006) 1040–1055.
[28] K. Miyazaki, K.W. Miyazaki, G. Matsumoto, Different representation of forthcoming reward in nucleus accumbens and medial prefrontal cortex, Neuroreport 15 (2004) 721–726.
[29] M. Shidara, B.J. Richmond, Anterior cingulate: single neuronal signals related to degree of reward expectancy, Science 296 (2002) 1709–1711.
[30] J.H. Sul, H. Kim, N. Huh, D. Lee, M.W. Jung, Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making, Neuron 66 (2010) 449–460.
[31] B.W. Balleine, A. Dickinson, The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety, Learn. Behav. 26 (1998) 46–59.
[32] L.H. Corbit, B.W. Balleine, The role of prelimbic cortex in instrumental conditioning, Behav. Brain Res. 146 (2003) 145–157.
[33] E. Coutureau, A.R. Marchand, G. Di Scala, Goal-directed responding is sensitive to lesions to the prelimbic cortex or basolateral nucleus of the amygdala but not to their disconnection, Behav. Neurosci. 123 (2009) 443–448.
[34] S.B. Floresco, A.E. Block, M.T. Tse, Inactivation of the medial prefrontal cortex of the rat impairs strategy set-shifting, but not reversal learning, using a novel, automated procedure, Behav. Brain Res. 190 (2008) 85–96.
[35] J.E. Haddon, S. Killcross, Prefrontal cortex lesions disrupt the contextual control of response conflict, J. Neurosci. 26 (2006) 2933–2940.
[36] J.D. Cohen, S.M. McClure, A.J. Yu, Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration, Philos. Trans. R. Soc. Lond. B: Biol. Sci. 362 (2007) 933–942.
[37] G. Aston-Jones, J.D. Cohen, An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance, Annu. Rev. Neurosci. 28 (2005) 403–450.
[38] H.S. Gompf, C. Mathai, P.M. Fuller, D.A. Wood, N.P. Pedersen, C.B. Saper, et al., Locus ceruleus and anterior cingulate cortex sustain wakefulness in a novel environment, J. Neurosci. 30 (2010) 14543–14551.
[39] B.F. Caracheo, E. Emberly, S. Hadizadeh, J.M. Hyman, J.K. Seamans, Abrupt changes in the patterns and complexity of anterior cingulate cortex activity when food is introduced into an environment, Front. Neurosci. 7 (2013) 74.
[40] N. Kolling, T.E. Behrens, R.B. Mars, M.F. Rushworth, Neural mechanisms of foraging, Science 336 (2012) 95–98.
[41] H.B.M. Uylings, H.J. Groenewegen, B. Kolb, Do rats have a prefrontal cortex, Behav. Brain Res. 146 (2003) 3–17.
[42] N.D. Daw, J.P. O'Doherty, P. Dayan, B. Seymour, R.J. Dolan, Cortical substrates for exploratory decisions in humans, Nature 441 (2006) 876–879.
[43] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, Boston, MA, 1998.
[44] N. Daw, Trial-by-trial data analysis using computational models, in: M.R. Delgado, E.A. Phelps, T.W. Robbins (Eds.), Decision Making, Affect, and Learning: Attention and Performance XXIII, England: Oxford University Press, Oxford, 2011, pp. 3–38.
[45] W. Schultz, Predictive reward signal of dopamine neurons, J. Neurophysiol. 80 (1998) 1–27.
[46] G. Paxinos, C. Watson, The Rat Brain, 6th ed., Academic Press, San Diego, CA, 2007.
[47] M.G. Packard, J.L. McGaugh, Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning, Neurobiol. Learn. Mem. 65 (1996) 65–72.
[48] N. Schmitzer-Torbert, A.D. Redish, Development of path stereotypy in a single day in rats on a multiple-T maze, Arch. Ital. Biol. 140 (2002) 295–301.
[49] A.M. Graybiel, Habits, rituals, and the evaluative brain, Annu. Rev. Neurosci. 31 (2008) 359–387.
[50] M.E. Ragozzino, S. Detrick, R.P. Kesner, Involvement of the prelimbic-infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning, J. Neurosci. 19 (1999) 4585–4594.
[51] V. Boulougouris, J.W. Dalley, T.W. Robbins, Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat, Behav. Brain Res. 179 (2007) 219–228.
[52] J.P. De Bruin, M.G. Feenstra, L.M. Broersen, M. Van Leeuwen, C. Arens, S. De Vries, et al., Role of the prefrontal cortex of the rat in learning and decision making: effects of transient inactivation, Prog. Brain Res. 126 (2000) 103–113.
[53] S. Kinoshita, C. Yokoyama, D. Masaki, T. Yamashita, H. Tsuchida, Y. Nakatomi, et al., Effects of rat medial prefrontal cortex lesions on olfactory serial reversal and delayed alternation tasks, Neurosci. Res. 60 (2008) 213–218.
[54] G. Schoenbaum, S.L. Nugent, M.P. Saddoris, B. Setlow, Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations, Neuroreport 13 (2002) 885–890.
[55] Y. Chudasama, T.W. Robbins, Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex, J. Neurosci. 23 (2003) 8771–8780.
[56] M.E. Ragozzino, J. Kim, D. Hassert, N. Minniti, C. Kiang, The contribution of the rat prelimbic-infralimbic areas to different forms of task switching, Behav. Neurosci. 117 (2003) 1054–1065.
[57] M.R. Stefani, K. Groth, B. Moghaddam, Glutamate receptors in the rat medial prefrontal cortex regulate set-shifting ability, Behav. Neurosci. 117 (2003) 728–737.
[58] S. Ghods-Sharifi, D.M. Haluk, S.B. Floresco, Differential effects of inactivation of the orbitofrontal cortex on strategy set-shifting and reversal learning, Neurobiol. Learn. Mem. 89 (2008) 567–573.
[59] H.H. Yin, B.J. Knowlton, The role of the basal ganglia in habit formation, Nat. Rev. Neurosci. 7 (2006) 464–476.
[60] H.H. Yin, S.B. Ostlund, B.J. Knowlton, B.W. Balleine, The role of the dorsomedial striatum in instrumental conditioning, Eur. J. Neurosci. 22 (2005) 513–523.
[61] P. Voorn, L.J. Vanderschuren, H.J. Groenewegen, T.W. Robbins, C.M. Pennartz, Putting a spin on the dorsal-ventral divide of the striatum, Trends Neurosci. 27 (2004) 468–474.
[62] P. Mailly, V. Aliane, H.J. Groenewegen, S.N. Haber, J.-M. Deniau, The rat prefrontostriatal system analyzed in 3D: evidence for multiple interacting functional units, J. Neurosci. 33 (2013) 5718–5727.
[63] N.K. Horst, M. Laubach, Reward-related activity in the medial prefrontal cortex is driven by consumption, Front. Neurosci. 7 (2013) 56.
[64] W.J. Kargo, B. Szatmary, D.A. Nitz, Adaptation of prefrontal cortical firing patterns and their fidelity to changes in action-reward contingencies, J. Neurosci. 27 (2007) 3548–3559.
[65] N. Insel, C.A. Barnes, Differential activation of fast-spiking and regular-Firing neuron populations during movement and reward in the dorsal medial frontal cortex, Cereb. Cortex 25 (2015) 2631–2647.
[66] B.G. Burton, V. Hok, E. Save, B. Poucet, Lesion of the ventral and intermediate hippocampus abolishes anticipatory activity in the medial prefrontal cortex of the rat, Behav. Brain Res. 199 (2009) 222–234.
[67] T.A. Stalnaker, M.R. Roesch, T.M. Franz, K.A. Burke, G. Schoenbaum, Abnormal associative encoding in orbitofrontal neurons in cocaine-experienced rats during decision-making, Eur. J. Neurosci. 24 (2006) 2643–2653.
[68] M.R. Roesch, Y. Takahashi, N. Gugsa, G.B. Bissonette, G. Schoenbaum, Previous cocaine exposure makes rats hypersensitive to both delay and reward magnitude, J. Neurosci. 27 (2007) 245–250.
[69] J.P. de Bruin, W.A. Swinkels, J.M. de Brabander, Response learning of rats in a Morris water maze: involvement of the medical prefrontal cortex, Behav. Brain Res. 85 (1997) 47–55.
[70] J.P. de Bruin, M.P. Moita, H.M. de Brabander, R.N. Joosten, Place and response learning of rats in a Morris water maze: differential effects of fimbria fornix and medial prefrontal cortex lesions, Neurobiol. Learn. Mem. 75 (2001) 164–178.
[71] R.J. McDonald, A.L. King, N. Foong, Z. Rizos, N.S. Hong, Neurotoxic lesions of the medial prefrontal cortex or medial striatum impair multiple-location place learning in the water task: evidence for neural structures with complementary roles in behavioural flexibility, Exp. Brain Res. 187 (2008) 419–427.
[72] M.E. Walton, D.M. Bannerman, M.F. Rushworth, The role of rat medial frontal cortex in effort-based decision making, J. Neurosci. 22 (2002) 10996–11003.
[73] D. Joel, I. Weiner, J. Feldon, Electrolytic lesions of the medial prefrontal cortex in rats disrupt performance on an analog of the Wisconsin Card Sorting Test, but do not disrupt latent inhibition: implications for animal models of schizophrenia, Behav. Brain Res. 85 (1997) 187–201.
[74] J.P. Johansen, H.L. Fields, B.H. Manning, The affective component of pain in rodents: direct evidence for a contribution of the anterior cingulate cortex, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 8077–8082.
[75] J.P. Johansen, H.L. Fields, Glutamatergic activation of anterior cingulate cortex produces an aversive teaching signal, Nat. Neurosci. 7 (2004) 398–403.
[76] I. Skelin, R. Hakstol, J. Vanoyen, D. Mudiayi, L.A. Molina, V. Holec, et al., Lesions of dorsal striatum eliminate lose-switch responding but not mixed-response strategies in rats, Eur. J. Neurosci. (2014).
[77] J.H. Sul, S. Jo, D. Lee, M.W. Jung, Role of rodent secondary motor cortex in value-based action selection, Nat. Neurosci. 14 (2011) 1202–1208.