PREFRONTAL REGULATION OF EXPLORATORY BEHAVIOR: COMPLEMENTARY FUNCTIONS OF MEDIAL PREFRONTAL AND ORBITOFRONTAL CORTICES IN A COMPETITIVE CHOICE TASK

BENJAMIN ARTHUR SEAN LIVINGSTONE
Bachelor of Science, University of Lethbridge, 2023

A thesis submitted
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

NEUROSCIENCE

Department of Neuroscience
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

Prefrontal Regulation of Exploratory Behavior: Complementary Functions of Medial Prefrontal and Orbitofrontal Cortices in a Competitive Choice Task

Ben Livingstone

Date of Defence: August 21, 2025

Dr. A. Gruber	Professor	Ph.D.
Dr. R. McDonald	Professor	Ph.D.
Thesis Co-Supervisors		
Dr. A. Luczak	Professor	Ph.D.
Thesis Examination Committee Member		
Dr. M. Tata	Professor	Ph.D.
Thesis Examination Committee Member		
Dr. S. Pellis	Professor	Ph.D.
Chair, Thesis Examination Committee		

Dedication

To my parents, siblings, friends, and labmates, who not only helped me immensely during this degree, but made it the time of my life. Thank you.

Abstract

Adaptive decision-making requires balancing exploration of uncertain options with exploitation of known rewards. This thesis investigates how the rat medial prefrontal cortex (mPFC) and orbitofrontal cortex (OFC) contribute to this trade-off during a competitive binary choice task (BCT), where stereotyped responses are punished and stochastic responding is rewarded. Male Long-Evans rats received excitotoxic lesions to the mPFC, OFC, or sham surgery, then performed the BCT across stable and volatile conditions. Behavior was analyzed using generalized linear mixed models to assess trial-level changes in extraneous feeder sampling (EFS), a non-rewarded exploratory action.

Lesions did not impair overall task engagement. However, OFC lesions caused a persistent elevation in EFS during the first post-operative session, consistent with impaired suppression of valueless behavior. In contrast, mPFC lesions led to an accelerated within-session decline in EFS, suggesting a deficit in sustaining exploratory strategies over time. Under environmental uncertainty, when the physical configuration of the task changed mid-session, both lesion groups showed a blunted increase in EFS and failed to reduce exploration across days.

These findings reveal a dissociation in stable environments: the OFC is critical for value-based suppression of exploration, while the mPFC supports its flexible maintenance. Under uncertainty, both regions are necessary for adaptive exploration, suggesting that they converge within a broader prefrontal network recruited during volatile conditions. This work clarifies the distinct and overlapping roles of the mPFC and OFC in regulating exploration and offers insight into prefrontal contributions to flexible decision-making.

Ethics Statement

Work described in this thesis received research ethics approval from the University of Lethbridge Animal Welfare Committee, Project Name “Prefrontal Contributions to Exploratory Behaviors”, No. 2212, 2024-05-20.

Use of Generative AI

Large-language models, such as ChatGPT and Gemini, were used to edit and proof-read this thesis.

Acknowledgements

I would like express my sincere gratitude to my co-supervisors Dr. Aaron Gruber and Dr. Robert McDonald. Their expert guidance and support were instrumental in the completion of this thesis. To my committee members, Dr. Matthew Tata and Dr. Artur Luczak, thank you for your critical perspectives that pushed this work to a higher standard.

As Dr. Gruber says, science is a team sport. This work would not have been possible without the dedicated work of Nhung Hong, Vicky Ivan, or Julia Medlicott. Thank you all so much for your help.

Finally, I wish to express my thanks to the funding agencies, Alberta Innovates and NSERC that provided generous financial support, which allowed me to dedicate my full attention to my studies and research. I am sincerely grateful for their contribution to my academic journey.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACC	Anterior Cingulate Cortex
ANOVA	Analysis of Variance
AP	Anteroposterior (stereotaxic axis)
A-O	Action-Outcome
BCT	Binary Choice Task
CI	Confidence Interval
DLS	Dorsolateral Striatum
DMS	Dorsomedial Striatum
DV	Dorsoventral (stereotaxis axis)
EFS	Extraneous Feeder Sampling
GLMM	Generalized Linear Mixed Model
IL	Infralimbic Cortex
IR	Infrared
ITI	Inter-Trial Interval
LED	Light-Emitting Diode
ML	Mediolateral (stereotaxic axis)
mPFC	Medial Prefrontal Cortex
NB	Negative Binomial
OFC	Orbitofrontal cortex
OR	Odds Ratio
PC	Personal Computer
PL	Prelimbic Cortex
RR	Rate Ratio
S-R	Stimulus-Response

Chapter 1: Introduction

Flexible behavior requires a trade-off between exploiting previously gained knowledge to reach a more desirable state and engaging in information-seeking behaviors to better understand the environment. This is known as the "explore-exploit dilemma" and is central to the fields of behavioral and computational reinforcement learning (Daw et al., 2006; Kaelbling et al., 1996; Krebs et al., 1978; Sutton & Barto, 1998). Reinforcement learning provides a powerful framework for understanding how organisms guide their decisions through learning from rewards and punishments. Foundational work by Sutton and Barto (Sutton & Barto, 1998) formalized reinforcement learning algorithms, introducing core concepts such as value functions, prediction error learning, the actor-critic framework, and exploration strategies like epsilon-greedy and softmax action selection. These computational principles have provided a basis for interpreting behaviour and neural signals in animals and humans alike. Complementing this theoretical foundation, Daw and colleagues (Daw et al., 2005, 2006) linked these reinforcement learning ideas to the functional anatomy of human decision-making, showing that distinct prefrontal regions contribute to the explore-exploit trade-off. Together, these contributions established the conceptual and experimental groundwork for studying exploration in neural systems. In this framework, an agent learns to assign values to different states or actions and uses these values to select actions that maximize future rewards. There has been considerable progress in this field with many experiments elucidating the biological mechanisms of reinforcement learning (Dayan & Niv, 2008; Lee et al., 2012; Rushworth & Behrens, 2008).

Studies on how animals manage the explore-exploit dilemma have implicated several brain regions and circuits. The midbrain dopamine system is thought to signal reward prediction errors that are crucial for the learning of reward values (Montague et al., 1996; Schultz et al., 1997). These dopaminergic signals are broadcast throughout the brain, but are particularly dense in the striatum, a key input nucleus of the basal ganglia (Weiner et al., 1991). Striatal subregions play distinct roles: the dorsomedial striatum (DMS) is implicated in goal-directed learning (Yin et al., 2005), whereas the dorsolateral striatum (DLS) is involved in habitual, stimulus-response learning (Featherstone & McDonald, 2004; McDonald & White, 1993; Yin et al., 2004). The nucleus accumbens in the ventral striatum is critical for motivation and processing reward value (Corbit & Balleine, 2011; Hart et al., 2014). Koralek & Costa, (2021) found that during periods of exploitation, dopamine neurons in the substantia nigra pars compacta (SNc) exhibited higher baseline activity, while locus coeruleus (LC) norepinephrine (NE) neurons showed reduced baseline activity. This complementary pattern suggests that dopamine promotes exploitative states, whereas LC–NE activity facilitates exploration. Consistent with this, (Tervo et al., 2014) demonstrated that LC–NE projections to the anterior cingulate cortex (ACC) drive a shift away from established strategies toward stochastic responding under conditions of uncertainty.. While many of these systems (excluding the largely sensorimotor-connected DLS) maintain strong interactions with prefrontal regions, the precise roles of subregions like the medial prefrontal and orbitofrontal cortices in orchestrating explore–exploit dynamics remain unclear.

The medial prefrontal cortex (mPFC) is comprised of three interconnected subregions: the anterior cingulate cortex (ACC), prelimbic cortex (PL), and infralimbic cortex (IL),

that have roles supporting the explore-exploit trade-off. The functions of these areas include monitoring performance, detecting errors, and updating action values to guide future choices (Euston et al., 2012; Miller & Cohen, 2001). Anatomically, the mPFC exhibits a dorsal-ventral gradient: dorsal PL and ACC circuits are heavily involved in goal-directed cognition and reward evaluation, while the ventral PL and IL modulate arousal, emotional responses, and extinction learning (Heidbreder & Groenewegen, 2003). This functional divergence reflects their afferent and efferent connections. The ACC and dorsal PL receive afferents from sensorimotor and association cortices, then send efferents to dorsal striatum, basolateral amygdala, thalamus and dorsal raphe (Gabbott et al., 2005; Hoover & Vertes, 2007). Ventral PL and IL receive stronger inputs from limbic regions such as the ventral hippocampus and amygdala (Hoover & Vertes, 2007). Ventral PL projects to nucleus accumbens, basolateral amygdala, and anterior hypothalamus, while IL sends outputs to the lateral septum, bed nucleus of the stria terminalis, multiple hypothalamic nuclei, central amygdala, and brainstem autonomic centres (Gabbott et al., 2005). All three subregions are innervated by neuromodulatory nuclei such as the ventral tegmental area and substantia nigra pars compacta (dopamine), locus coeruleus (norepinephrine), and dorsal raphe (serotonin) (Hoover & Vertes, 2007). These neuromodulatory systems are known to affect behaviour by adjust neural processing in response to salient contextual variables. This circuitry indicates that the mPFC is well placed to support the cognitive functions used in decision making.

Recent work has provided more direct evidence for the mPFCs role in exploration. One group found that ACC activity was suppressed in a behavioral test that required stochastic responses (Tervo et al., 2014). Another study by the same group (Tervo et al., 2021)

found distinct subcircuits within the ACC that were individually involved in exploitative or exploratory actions. These studies suggest both that the ACC plays a role in action selection and strategy arbitration, and that there must be mechanisms to override the ACC when its predictions are not successful. An electrophysiological study found that transitions from exploration to exploitation phases of a behavioral task led to abrupt changes in ACC activity, again suggesting that the ACC has multiple circuits responsible for strategy arbitration (Caracheo et al., 2013). Furthermore, a recent study found a specific role for the ACC, but not the orbitofrontal cortex, in choices involving a deliberate search for information (González et al., 2024). Another study found that the mPFC was involved in inhibiting previously learned strategies in favour of a new one (Ragozzino, 2007). Taken together, these findings suggest that the mPFC is not just involved in general performance monitoring but has a more direct role in supporting exploratory actions—potentially by overriding prepotent, exploitative choices.

To our knowledge, there have been no studies directly investigating how exploratory behaviors are moderated by the prelimbic (PL) and infralimbic (IL), the other main subdivisions of the mPFC, in a competitive environment. However, their specific roles in the explore-exploit trade-off can be inferred from other decision-making studies. It is well-supported that there is a functional dichotomy between the PL and IL. The PL is considered essential for encoding and executing goal-directed, action-outcome (A-O) associations, which allows for flexible, adaptive behavior when circumstances change (Balleine & Dickinson, 1998). In contrast, the IL is more critically involved in the expression of well-learned, stimulus-response (S-R) habits, particularly in inhibiting previously learned goal-directed actions (Killcross & Coutureau, 2003). This division is

crucial in the context of exploration and exploitation; goal-directed control by the PL is paramount for exploring new strategies when established ones are no longer optimal, while the IL is thought to facilitate the exploitation of learned, habitual responses.

The role of the PL in promoting behavioral flexibility is particularly relevant to our study. In tasks that require animals to override a prepotent response in favor of a more adaptive one, such as the attentional set-shifting paradigm, the PL is critical for the "extra-dimensional shift" where the animal must attend to a previously irrelevant stimulus dimension (Birrell & Brown, 2000; Ragozzino, 2007). Furthermore, in the Rat Gambling Task, which models complex decision-making under uncertainty, inactivation of the PL impairs the ability to develop and maintain an optimal, reward-maximizing strategy (Zeeb et al., 2015). These findings suggest the PL uses higher-order information about task rules and environmental uncertainty to guide action selection (Euston et al., 2012; Miller & Cohen, 2001). This form of “top-down” control is vital in our competitive task, as the opponent's algorithm is designed to learn and punish predictable behavioral patterns. Consequently, simple choice strategies are ineffective, and success requires the animal to suppress prepotent response tendencies in favor of stochastic, unpredictable patterns of behavior. Therefore, we hypothesize that the PL plays a key role in managing the explore-exploit trade-off. Specifically, the PL may be necessary for sustaining exploration when uncertainty is high, and a shift away from a currently failing exploitative strategy is required. We expect that lesions of the PL will compromise the expression of exploratory behaviors needed to adapt to complex environments.

The orbitofrontal cortex (OFC) is another prefrontal region with a well-established role in decision-making, particularly in representing the value of expected outcomes (Izquierdo,

2017; Schoenbaum et al., 1998). Recent work suggests that the OFC furnishes a cognitive map of the task "state-space"—the set of all distinct situations an animal might encounter during the task (Wilson et al., 2014). A state is a snapshot of behaviorally relevant information: external cues (lights on/off), recent outcomes, and latent or internal variables such as motivational state. By keeping track of which state it is in, the animal can attach the correct value to each available action and switch strategies when the state changes. We hypothesize that the OFC's role in the explore-exploit dilemma is to represent and update outcome expectancies through this state-space representation. This is supported by studies showing that OFC lesions impair reversal learning—where an animal must shift its choice from a previously rewarded option to a newly rewarded one (Schoenbaum et al., 2002; Tait & Brown, 2007). The impairment in reversal learning may be caused by a loss of ability to devalue the previously correct strategy. More recent evidence suggests the OFC's contribution is dynamic; it is particularly crucial for adaptive behavior when reward values are unstable (Riceberg & Shapiro, 2012). Another recent study showed that neural activity in the primate OFC encodes information about both the value of the current choice, and the uncertainty associated with alternative choices (Costa & Averbeck, 2020), providing information relevant to the explore-exploit trade-off. Furthermore, projections from the OFC to secondary motor cortex have been shown to mediate the exploitation of learned rules, suggesting a direct role in implementing value-based choices (Schreiner & Gremel, 2018). This converging evidence supports our hypothesis that the OFC contributes to the explore-exploit dilemma by representing state-space and mitigates expression of zero-value behaviors.

To investigate the distinct roles of the mPFC and OFC in decision-making, the present work used a rat binary choice task. In this task, a competitive algorithm attempts to predict the rat's next choice, thereby punishing stereotyped behaviors and rewarding a stochastic, unpredictable strategy (Gruber et al., 2017; Ivan et al., 2018). Like a soccer player during a penalty kick, if the player always kicks towards the same spot in the net, the goaltender can exploit this pattern. This dynamic environment is ideal for dissociating the functions of the mPFC and OFC. Specifically, we focused on how excitotoxic lesions to these regions would affect spontaneous, non-rewarded exploratory behavior. In this paradigm, this is demonstrated by extraneous feeder sampling (EFS), a term used to describe the post-choice exploration of the other feeder. We hypothesized that the OFC—critical for representing action-outcome values—is necessary to suppress this zero-value behavior, therefore OFC lesions lead to a persistent increase in EFS. Additionally, we predicted that the mPFC influences how EFS is modulated over time due to its role in the top-down regulation of strategic behavior. This study investigated how these two key prefrontal regions contribute to the regulation of exploratory behavior through a well-established competitive task.

Chapter 2: Materials and Methods

All procedures were approved by the University of Lethbridge Animal Welfare Committee, in accordance with the Canadian Council of Animal Care Guidelines.

2.1 Subjects

45 male Long-Evans rats (Charles River. QC, Canada; 350-512g) were pair or triplet housed in transparent plastic cages with corncob bedding and a black PVC tube for enrichment. Animals were handled daily for one week before training began. Animals had

free access to water for 30 minutes a day after testing and *ad libitum* access to water 4 days prior to surgery and for 10 days post-surgery. Animals had *ad libitum* access to food for the entirety of the experiment. Of the 45 animals, 8 did not meet the training criteria, and 12 did not meet the lesion criteria.

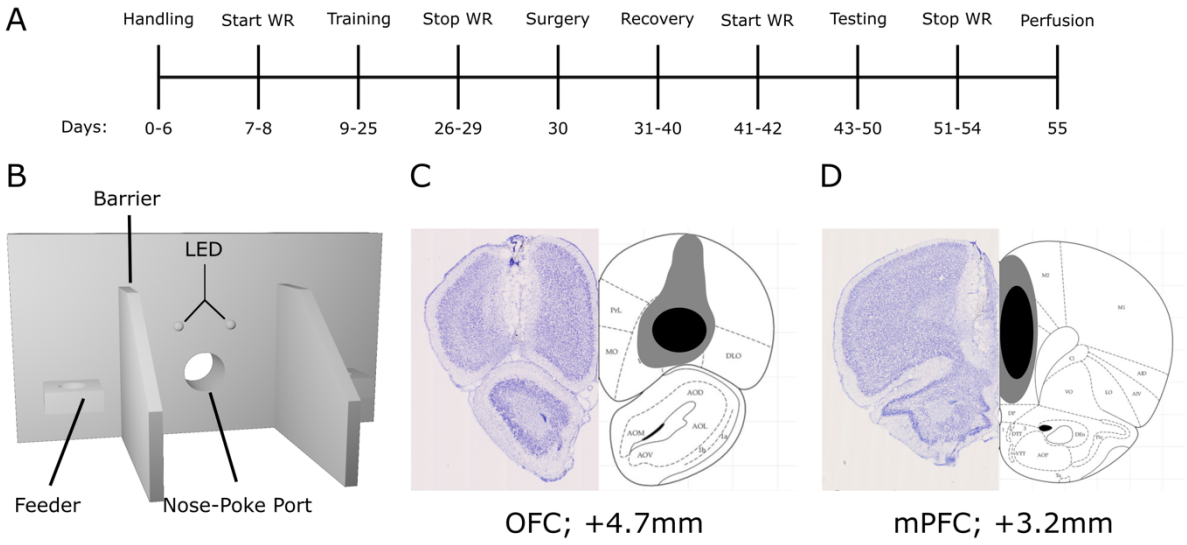

Figure 1: Overview of the scientific design. A: Experimental Timeline. B: Model of behavioral apparatus. Nose-poke port with IR sensors was used to start a trial. Animal then locomoted around the barriers to either feeder, which dispensed a sucrose reward. C: Sample OFC lesion; Nissl (Cresyl Violet) stain on left side, annotated atlas section with minimal (black) and maximal (grey) lesion extents. D: Sample mPFC lesion; Nissl stain on the left, minimal and maximal lesions on the right.

2.2 Behavioral Task

This behavioral task is a binary choice task (BCT) that took place in one of six identical custom built 26x26cm aluminum boxes. One wall of the box had a nose-poke port flanked by liquid reward dispensing feeders on either side (Fig. 1.B). The nose-poke port was outfitted with infrared beam-break sensors to detect when the animal entered. The feeders were outfitted with two sets of beam break sensors. One to detect when the animal entered the feeder, and another to detect licks. Above the nose-poke port were two

LED lights, which signaled the animal to start a trial by entering the nose-poke port. When a trial was started, a tone would play to cue the animal to move to a feeder. Between the nose-poke port and feeders was a 13cm barrier which was used to introduce a choice cost and reduce side bias. If the trial was rewarded, 60μL of a 10% sucrose solution was dispensed. If the trial was not rewarded, a white LED illuminated the box to signal a loss. Control of the behavioral task was automated with an Arduino Mega microcontroller. The Arduino sent serial data to an external PC running custom software.

Training was achieved by gradually shaping components of the desired behavior. In the initial shaping stage, there were no barriers, and animals were rewarded every time they entered the nose-poke port then entered a feeder within 15s. Once the animal completed 150 trials, the session was terminated. Animals were then switched to another reward protocol in which trials were rewarded randomly, with a mean reward rate of 50%. The following sessions used a competitive algorithm that would learn from the animals' previous decisions and reward the well the algorithm deems the animal is less likely to enter. This algorithm is detailed below. Aluminum barriers were gradually introduced, starting with a 4cm barrier, followed by 8cm and finally 13cm barriers. The training criteria for animal inclusion were to complete 150 trials within 45 minutes with the 13cm barrier.

Once training criteria were met, animals underwent three days of baseline testing, each consisting of a 45-minute session with the 13 cm barrier, during which there was no limit on the number of trials. After the animals had recovered from surgery, they underwent three days of testing identical to the baseline testing (Fig. 1.A).

The competitive algorithm used in this experiment was identical to those in past Gruber lab studies (Gruber et al., 2017; Gruber & Thapa, 2016; Skelin et al., 2014) and was adapted from algorithm 2 in (Lee et al., 2004). This algorithm used the entire choice and reward history of the animal in a session to predict the animals next choice. Conditional probabilities of choosing either feeder, given the animals' choices and rewards on the previous 4 trials, were calculated. These conditional probabilities were tested against the null hypothesis that each probability was 0.5. If the null hypothesis was rejected via a binomial test, the largest probability feeder was set as unrewarded. If the null hypothesis was not rejected, the algorithm rewarded a random well with 0.5 probability. In this game, the Nash equilibrium is reached when the animal adopts a mixed strategy by randomizing its choices. Thus, an ideally performing animal should receive rewards 50% of trials.

After the post-surgery tests, a variation of the BCT was used to assess the animal's response to changes in the environment. This variation included swapping the length of the barrier separating the nose-poke port from the feeders. The session began with the 13cm barrier, then after 60 trials this was swapped with either a shorter 8cm barrier, a longer 20cm barrier, or a control 13cm barrier.

2.3 Surgery & Lesion Verification

Animals that completed training were randomly assigned to receive either a mPFC lesion, OFC lesion, or a sham lesion. Animals had *ad libitum* access to water for four days prior to surgery. On the day of the surgery, animals were treated with 0.03mg/kg Buprenorphine subcutaneously 30 minutes prior to the induction of anesthesia via isoflurane. Excitotoxic lesions were performed at the following coordinates from bregma,

in mm (AP, ±ML, DV): OFC (4, 2, -4.2) (3.4, 2.7, -5.4), mPFC (3.5, 0.7, -5). NMDA (7.5μg/μL) was infused at 0.1μL/minute over 4 minutes for OFC lesions and 0.2μL/minute over 3 minutes for mPFC lesions. The needle remained in situ for 2 minutes to permit diffusion in both lesion groups. Animals were given 1mg/kg Inflacam (meloxicam) subcutaneously immediately after surgery to minimize discomfort. Animals were monitored at least twice daily for a minimum of 4 days after surgery to track recovery. Animals were given 10 days to recover before behavioral tests resumed

After testing was completed, animals were perfused and their brains collected and immersed in a paraformaldehyde solution and stored at 4°C. The brains were kept in the paraformaldehyde solution for 24 hours and were then transferred to a 30% sucrose and 0.03% sodium azide solution. The brains were sectioned into 30μm slices and mounted on gelatin-coated slides. The sections were then Nissl stained, and cover slipped. The slides were imaged using a NanoZoomer slide scanner (Hamamatsu, Shizuoka Japan). For each lesioned brain, the scans were compared to a rat brain atlas (Paxinos & Watson, 2006) to determine the lesion extent. For each slice in the brain atlas, corresponding minimum and maximum lesions were drawn. Animals with significant damage to surrounding areas were excluded from the analysis (Fig. 1.C-D).

2.4 Behavioral Analysis

Behavioral data was analyzed using MATLAB (version 2022B) and R (version 4.4.3). Several behavioral measures were quantified in this BCT. Licking behavior was quantified both before a reward was delivered as well as total licks in each trial. This was to measure motivational and consummatory behaviors. Response time was calculated as

the time between leaving the nose-poke port and entering a feeder. Likewise, inter-trial interval (ITI) was calculated as the time between leaving the feeder and entering the nose-poke port. Extraneous feeder sampling (EFS) was a binary value that was true if the animal entered both feeders during a trial and false otherwise. Behavioral data were analyzed using the Kruskal-Wallis test and generalized linear mixed models (GLMM). The Kruskal-Wallis test was used to analyze group differences in several measures of task performance (e.g. number of trials, percent of correct trials, reaction time, response entropy). The Kruskal-Wallis test was used over traditional one-way ANOVA because the data were not normally distributed and heteroscedastic. GLMMs were used to analyze behaviors with repeated measures within subjects and across trials. These measures include EFS, ITI, licks before reward, and total licks per trial. GLMM was used over repeated measures ANOVA as GLMMs are not subject to the assumptions of normality and sphericity.

2.5 Generalized Linear Mixed Models (GLMM)

Repeated behavioral observations were analysed with generalised linear mixed models (GLMMs), which extend generalised linear models by adding random effects that account for the hierarchical structure—and therefore the non-independence—of the data (Bolker et al., 2009). Separate models were fitted for each outcome: extraneous feeder sampling (EFS; binary), inter-trial interval (ITI; positive, right-skewed continuous), and lick count (counts). Error–link pairs were chosen to match the data type (binomial–logit for EFS, log-normal with identity link for ITI, negative binomial [nbiom2]-log for licks). Surgical group (sham, mPFC lesion, OFC lesion) was specified as a fixed effect, while rat identity was included as a random intercept. Models were fit using the R Package

glmmTMB version 1.1.11 (Brooks et al., 2017; McGillycuddy et al., 2025). This package uses the Laplace approximation and automatic differentiation for maximum-likelihood estimation. Model residuals were inspected with DHARMa v 0.4.7 (Hartig, 2024) to verify that distributional assumptions and dispersion were appropriate.

Chapter 3: Results

We trained 45 rats in a matching-pennies style binary choice task. Of the 45 rats, 8 failed to meet the learning criteria for the BCT and did not undergo surgery. 14 animals received OFC lesions. Due to involvement of adjacent areas such as secondary motor cortex, 8 animals were excluded from analysis. 14 animals received mPFC lesions. Due to incomplete destruction of the region encompassing the ACC, PL, and IL, 4 animals were excluded. We examined 25 rats, 9 of which received sham lesions, 6 received OFC lesions, and 10 mPFC lesions. An additional 3 days of the BCT were run where the barrier separating the nose-poke port from the feeders was removed and replaced with one of shorter, longer, or equal length midway through the session to assess how lesion condition affected responses to changing environments.

3.1 Baseline task engagement

To determine whether surgical treatment impaired general task performance, we compared four session-level metrics across groups: total trials completed, proportion of correct choices, median reaction time, and response entropy. Kruskal–Wallis H tests showed no significant group differences for any measure, trials completed ($p = 0.279$), correct-choice proportion ($p = 0.731$), reaction time ($p = 0.269$), or response entropy ($p = 0.705$; Fig. 2A–D). Because none of these global performance indices differed by group,

any effects of the lesions described below are unlikely to be attributed to overall motivational, accuracy-related, or timing deficits.

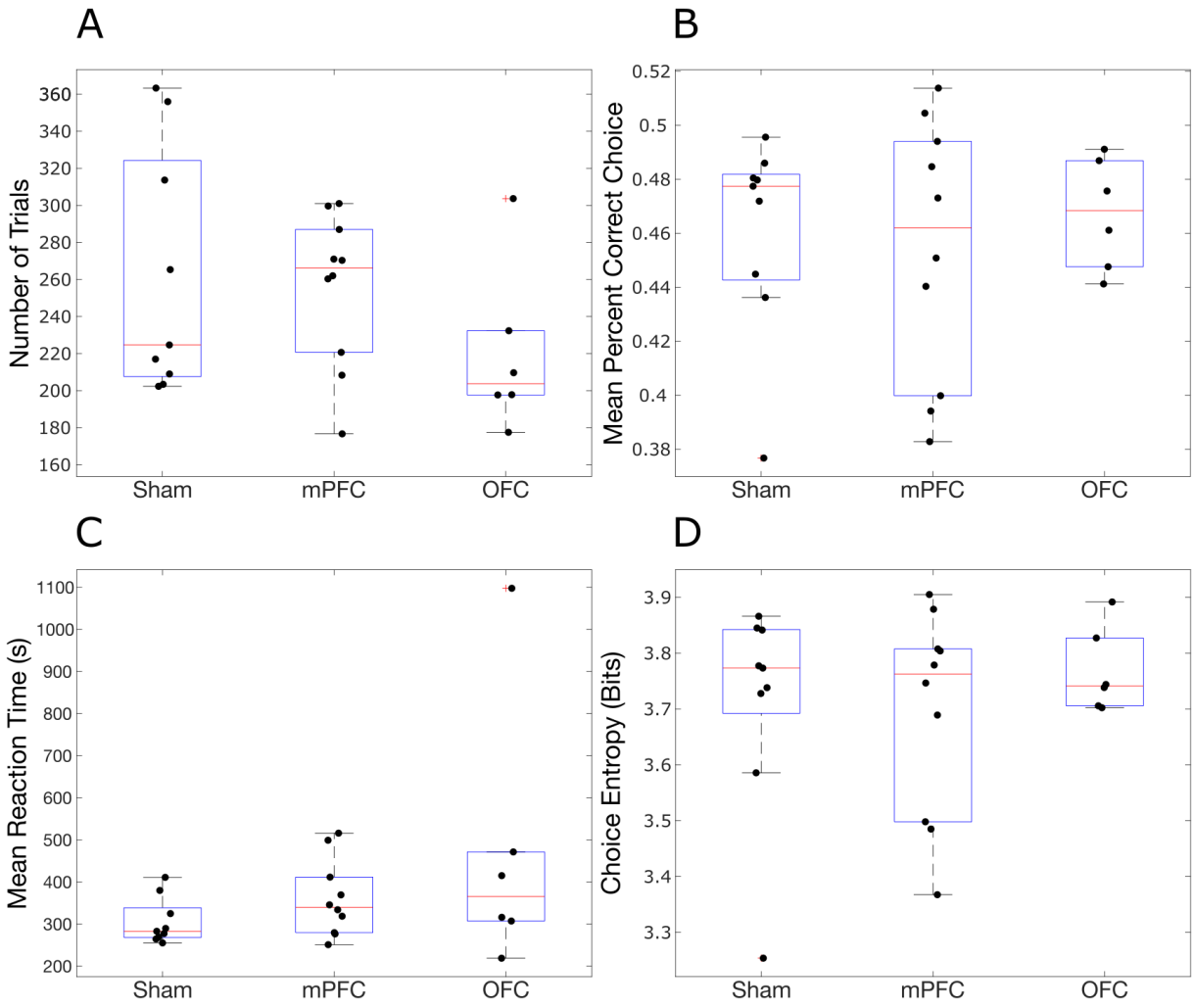

Figure 2: Box plots of task performance measures for each animal, averaged across sessions. A: Number of trials completed. B: Proportion of correct choices. C: Reaction time. D: Shannon Entropy of choices.

3.2 General analytic strategy

All inferential statistics were obtained from generalized-linear mixed-effects models (GLMMs) fitted to single-trial data. Random intercepts for individual rats captured

repeated-measures structure; fixed-effects structures are (Variable ~ Trial Number * Group ID + (1 | Rat ID)) unless otherwise specified. Plots shown in the manuscript are for visualization only and were generated from binned data; the statistics reported here were computed on the full, un-binned trial set. Each link function produces different estimates, reported here are the exponentiated estimates. The identity link function in the log-normal model used to analyze inter-trial interval times produces a multiplicative factor ratio (shortened to "ratio" for this paper). Logit link functions, used to analyze binary extraneous feeder sampling data, produce odds ratio (OR) estimates, and log link functions, used to analyze count data for licking measurements, produce rate ratios (RR).

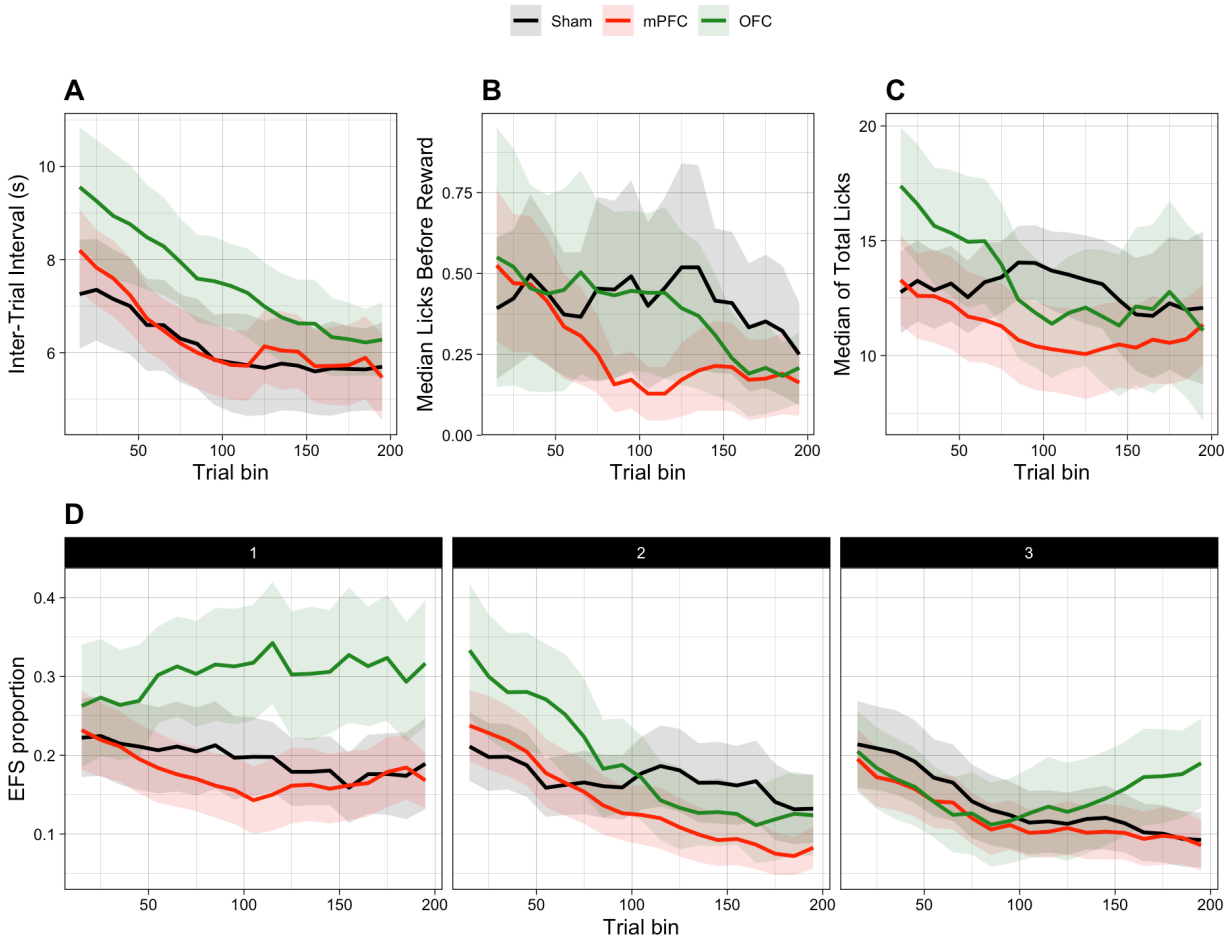

Figure 3: Within-session dynamics of behavioral variables, the shaded region indicates the standard error. A: Mean Inter-trial interval. B: Number of licks total that the animal performed in each trial. C: Number of times the animal licked the feeder before the reward was given. D: Probability of performing extraneous feeder sampling across 3 sessions.

Table 1: Results of GLMM analysis on behavioral measures

component		EFS (binomial)	Licks Before Reward (NB)	Licks Total (NB)	log(ITI) (Gaussian)
conditional	Intercept	0.442 [0.316, 0.619]	2.676 [1.951, 3.671]	18.585 [15.660, 22.057]	7751.295 [6660.464, 9020.780]
		p = <0.001***	p = <0.001***	p = <0.001***	p = <0.001***

component		EFS (binomial)	Licks Before Reward (NB)	Licks Total (NB)	log(ITI) (Gaussian)
	mPFC vs Sham	1.150 [0.720, 1.838]	0.868 [0.562, 1.341]	0.873 [0.689, 1.105]	1.004 [0.814, 1.237]
		p = 0.559	p = 0.522	p = 0.258	p = 0.972
	OFC vs Sham	2.031 [1.205, 3.424]	1.062 [0.646, 1.746]	1.198 [0.914, 1.570]	1.196 [0.941, 1.520]
		p = 0.008**	p = 0.811	p = 0.192	p = 0.144
	Trial Index	0.995 [0.993, 0.997]	0.997 [0.995, 0.999]	0.999 [0.999, 1.000]	0.997 [0.996, 0.998]
		p = <0.001***	p = 0.001**	p = 0.166	p = <0.001***
	Session	0.862 [0.781, 0.951]			
		p = 0.003**			
	mPFC × Trial	0.996 [0.993, 0.998]	0.999 [0.996, 1.001]	0.999 [0.998, 1.000]	1.000 [0.999, 1.000]
		p = 0.002**	p = 0.246	p = 0.086+	p = 0.397
	OFC × Trial	1.000 [0.997, 1.003]	1.001 [0.998, 1.003]	0.999 [0.997, 1.000]	1.000 [0.999, 1.001]
		p = 0.923	p = 0.686	p = 0.084+	p = 0.854
	mPFC × Session	1.007 [0.876, 1.158]			
		p = 0.919			
	OFC × Session	0.780 [0.670, 0.909]			
		p = 0.001**			
dispersion	Intercept		1.263 [1.250, 1.277]	2.551 [2.484, 2.622]	

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

3.3 Inter-trial interval (ITI)

It has been shown in previous studies that changes in ITI can alter choice strategy (Gruber & Thapa, 2016). Thus, it was important to determine if lesion condition affected ITI times. A log-normal GLMM (identity link function) revealed a strong effect of trial progression (ratio = 0.997, 95% CI [0.996, 0.998], $p < 0.001$), similar to the within-session ITI after loss decrease described by (Gruber & Thapa, 2016). Neither mPFC (ratio = 1.004, 95% CI [0.814, 1.237], $p = 0.972$) nor OFC (ratio = 1.196, 95% CI [0.941, 1.520], $p = 0.144$) main effects reached significance, and there were no significant interactions between trial number and mPFC (ratio = 1.0, 95% CI [0.999, 1.000], $p = 0.397$) or OFC (ratio = 1.0, 95% CI [0.999, 1.0], $p = 0.854$). This shows that, both lesion groups shortened their ITIs across the session at the same rate as shams (Fig. 3.A), indicating that lesion condition did not affect ITI in the BCT. As mentioned above, ITI length correlates with choice strategy.

3.4 Extraneous feeder sampling (EFS)

We next sought to determine if lesions of the mPFC or OFC affected rates of EFS, an off-task exploratory behavior. Previous studies have shown that EFS decreases both within-session and between-sessions (Gruber et al., 2017). To formally model these trends, we fit a GLMM (binomially distributed with logit link function) with fixed effects for lesion group (Sham, mPFC, OFC), trial number within the session, session number, and their interactions with group (EFS ~ Trial Number * Group ID + Session Number * Group ID + (1 | Animal ID)). This model revealed a significant main effect of trial number (OR = 0.995, 95% CI [0.993, 0.997], $p < 0.001$) suggesting EFS decreases within session. There

was also a main effect of session number (OR = 0.862, 95% CI [0.781, 0.951], $p = 0.003$) suggesting that EFS declines between sessions as well as within sessions (Fig. 3D). Interestingly, there was also a main effect of OFC lesion (OR = 2.031, 95% CI [1.205, 3.424], $p = 0.008$) suggesting that OFC lesioned animals had significantly higher levels of EFS relative to shams during the initial session. There were also significant interactions in this model, including mPFC lesion × trial number (OR = 0.996, 95% CI [0.993, 0.998], $p = 0.002$) showing that mPFC lesioned animals decreased EFS behavior within-sessions faster than shams. We also saw a significant OFC lesion × session number interaction (OR = 0.780, 95% CI [0.670, 0.909], $p = 0.001$) showing that OFC lesioned animals decreased EFS behavior between-sessions faster than sham. Taken together these results show that lesions to the prefrontal cortex significantly affect exploratory behaviors.

3.5 Licks before reward

We then wanted to determine if lesion condition influenced motivational behaviors. The first we examined was the number of licks the animal performed before the reward was delivered. Pre-reward licking decreased with Trial (negative binomial GLMM, log link; RR = 0.997, 95% CI [0.995, 0.999], $p = 0.001$) showing that this anticipatory measure decreased as the session progressed (Fig. 3.C). Neither the mPFC nor OFC groups had significant Group × Trial interactions (mPFC × Trial: RR = 0.999, 95% CI [0.996, 1.001] $p = 0.246$. OFC × Trial: RR = 1.001, 95% CI [0.998, 1.003], $p = 0.686$; Fig 3B). We found no evidence that the change in anticipatory licking over trials differed between the lesion and sham groups.

3.6 Total licks per trial

We also sought to determine if total licks per trial differed between the lesion groups. Total licking showed no significant main effect of Trial (RR = 0.999, 95% CI [0.999, 1.0], $p = 0.166$) and no main effects of group (mPFC: RR = 0.873, 95% CI [0.689, 1.105], $p = 0.258$; OFC: RR = 1.198, 95% CI [0.914, 1.570], $p = 0.192$) (Fig. 3C). Group × Trial interactions were also not significant. (mPFC × Trial: RR = 0.999, 95% CI [0.998, 1.0], $p = 0.086$; OFC × Trial: RR = 0.999, 95% CI [0.997, 1.0], $p = 0.84$; Fig 3D). The absence of significant main effects and interactions shows that trial number and lesion condition did not significantly affect licking in this paradigm.

3.7 Swapping barriers

Lastly, we sought to determine how changing the length of the barriers separating the nose-poke port from the feeders affected EFS. This change alters the configuration of the chamber and is thought to increase uncertainty about the task. Previous studies found that swapping barriers led to increased EFS behaviors when the hallway was swapped to shorter and longer barriers compared to shams (Gruber et al., 2017). We wanted to determine if lesion group influenced this increase in EFS. To analyze this, we used a binomially distributed GLMM as with the EFS analysis. We added a term for the trials to indicated if they occurred pre- or post-barrier swap. This analysis revealed no significant main effect of lesion group (mPFC: OR = 0.703, 95% CI [0.415, 1.192], $p = 0.191$; OFC: OR = 0.682, 95% CI [0.373, 1.248], $p = 0.215$) this indicates that the average EFS across sessions was not significantly different between groups. There was a significant main effect of trial number (OR = 0.991, 95% CI [0.987, 0.996], $p < 0.001$) showing again that EFS behavior decreases within the session (Fig. 4.A). Likewise, there was a significant

main effect of Session (OR = 0.700, 95% CI [0.606, 0.808], $p < 0.001$) (Fig. 4.B). Finally, there was a significant main effect of barrier swap (OR = 2.215, 95% CI [1.523, 3.222], $p < 0.001$), showing that swapping barriers significantly increased EFS behaviors. Interestingly, in this paradigm, both mPFC and OFC lesion groups showed significant interactions with session number (mPFC × Session Number: OR = 1.422, 95% CI [1.175, 1.721], $p < 0.001$; OFC × Session Number: OR = 1.470, 95% CI [1.162, 1.861], $p = 0.001$) This shows that, unlike sham animals, lesioned animals failed to decrease EFS between sessions. This contrasts with the results without the barrier swap in which only the OFC × Session Number interaction was significant. There were also significant interactions between lesion group and barrier swap (mPFC × Barrier Swap: OR = 0.619, 95% CI [0.385, 0.995], $p = 0.048$; OFC × Barrier Swap: OR = 0.425, 95% CI [0.250, 0.724], $p = 0.002$). This shows that animals with mPFC or OFC lesions had a significantly blunted exploratory response to the barrier swap compared to sham animals.

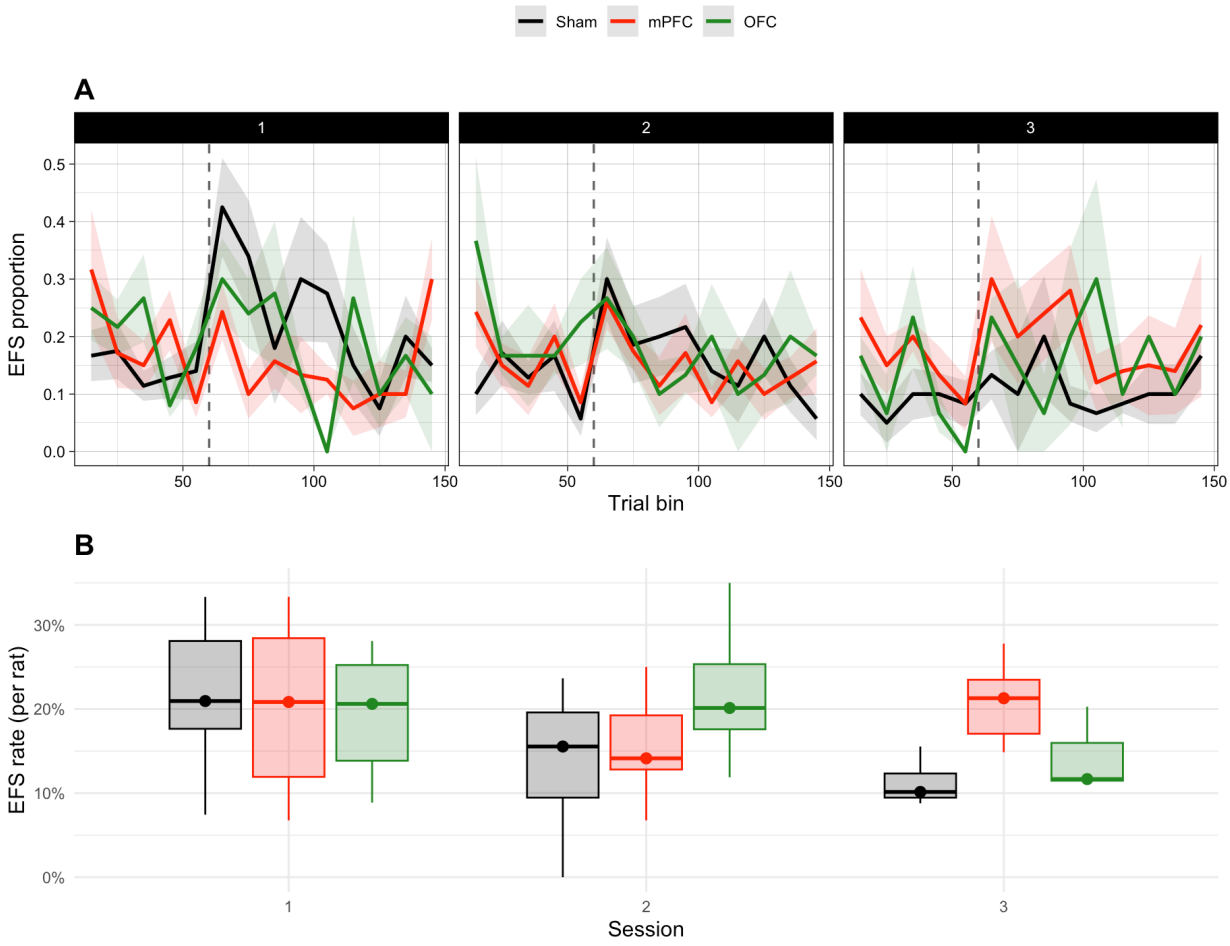

Figure 4: Results of barrier swapping. A: Shows within-session proportion of EFS in each session. Trials are binned in groups of 10 trials. Bins 1-15 are session 1, 16-30 are session 2, and 31-45 are session 3. Barrier swaps occurred on trial 60 in each session. B: Shows between-session performance only. Each point is averaged over the entire session.

Table 2: Results of barrier swapping GLMM analysis

component		EFS - Barrier Swap (binomial)
conditional	Intercept	0.431 [0.295, 0.631]
		p = <0.001***
	mPFC vs Sham	0.703 [0.415, 1.192]
		p = 0.191

component		EFS - Barrier Swap (binomial)
	OFC vs Sham	0.682 [0.373, 1.248]
		p = 0.215
	Trial Index	0.991 [0.987, 0.996]
		p = <0.001***
	Session	0.700 [0.606, 0.808]
		p = <0.001***
	Barrier Swapped	2.215 [1.523, 3.222]
		p = <0.001***
	mPFC × Trial	1.001 [0.996, 1.007]
		p = 0.606
	OFC × Trial	1.005 [0.999, 1.011]
		p = 0.118
	mPFC × Session	1.422 [1.175, 1.721]
		p = <0.001***
	OFC × Session	1.470 [1.162, 1.861]
		p = 0.001**
	mPFC × Barrier Swap	0.619 [0.385, 0.995]
		p = 0.048*
	OFC × Barrier Swap	0.425 [0.250, 0.724]
		p = 0.002**

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Chapter 4: Discussion

4.1 Summary of Findings

This study investigated the roles of the medial prefrontal cortex (mPFC) and orbitofrontal cortex (OFC) in regulating off-task exploratory behavior in a competitive binary choice task. We found a dissociation between the functions of these two regions. Excitotoxic lesions of the mPFC resulted in an accelerated within-session decline of extraneous feeder sampling (EFS), a spontaneous exploratory behavior. In contrast, OFC lesions led to a sustained elevation of EFS throughout the first post-operative session, which was followed by a more rapid between-session decline compared to sham controls. These specific effects on exploration were not likely attributable to general performance deficits, as there were no group differences in trials completed, choice accuracy, or reaction time.

Furthermore, convergent effects were observed in a barrier-swapping manipulation, as both lesion groups exhibited two key changes relative to shams: they failed to suppress EFS across sessions and they showed a blunted exploratory response to the barrier swap.

4.2 Role of mPFC in Competitive Choice Tasks

Our primary finding regarding the mPFC was that lesioned animals exhibited a significantly faster within-session decrease in EFS compared to sham controls. This suggests that while the mPFC is not necessary for initiating exploratory actions, it is critical for sustaining them over the course of a behavioral session. In our competitive task, where the optimal strategy is unpredictable responding, animals must allocate cognitive resources to the demanding on-task goal of generating variable choices while

also deciding whether to engage in off-task exploratory behaviors. The inability of our mPFC-lesioned animals to maintain EFS may reflect a failure in this high-level strategy arbitration.

This interpretation aligns with the known functions of key subregions within the mPFC, namely the anterior cingulate (ACC) and prelimbic (PL) cortices. The ACC is thought to be important for directing the exploration of alternative strategies (Tervo et al., 2021), potentially providing the drive to seek information. The PL cortex complements this function by being important for the flexible selection of actions (Capuzzo & Floresco, 2020), encoding task rules (Rich & Shapiro, 2009), and maintaining active, goal-directed behaviors (Gourley & Taylor, 2016). In essence, the ACC may signal the value of exploring, while the PL helps organize and sustain that exploration as a coherent behavioral strategy. Therefore, the accelerated decline in EFS in our mPFC-lesioned animals likely reflects the simultaneous disruption of these complementary functions, leading to an inability to maintain the active, exploratory component of the animal's behavioral repertoire in the face of ongoing task demands.

4.3 Role of OFC in Competitive Choice Tasks

In contrast to the mPFC group, animals with OFC lesions exhibited significantly higher rates of extraneous feeder sampling (EFS) during their first post-operative session compared to sham controls. Rather than interpreting this as the animals treating the familiar environment as entirely novel, a more parsimonious explanation is that the OFC is critical for rapidly applying learned action-outcome values in a specific context. In this well-learned task, the exploratory action of EFS has a learned value of zero. The OFC, which is thought to represent a cognitive map of the task state-space that links context to

value (Wilson et al., 2014), normally acts to suppress such inefficient, zero-value behaviors from the outset. The lesion-induced elevation in EFS during the first session is consistent with the OFC's role in uncertainty and volatility encoding. Specifically, the damage to the OFC may have degraded the stable contextual representation of the task's rules, functionally mimicking an increase in environmental volatility or behavioural uncertainty that the animals resolve through increased exploration (EFS). This aligns with findings that OFC projections are crucial for exploiting learned rules (Schreiner & Gremel, 2018), and that without the OFC, animals are less able to use expected outcomes to guide their choices.

Interestingly, the elevated EFS in the OFC group was not sustained. We observed a significant OFC lesion × session number interaction, indicating that these animals suppressed their EFS behavior more rapidly across the three days of testing than sham animals. This dynamic recovery suggests that while the OFC is critical for the initial, value-based deployment of a context-appropriate strategy, its absence allows slower, non-OFC-dependent learning mechanisms to compensate. A likely candidate for this compensatory learning is the dorsolateral striatum (DLS) habit system. The DLS is known to be essential for the incremental formation of stimulus-response (S-R) habits that are less dependent on rich outcome representations (Yin et al., 2004). While this interpretation remains speculative, it provides a neurobiologically plausible explanation for our data. Over several sessions, the repeated, non-reinforced nature of EFS may lead to its gradual suppression via instrumental extinction, a form of S-R learning often attributed to the DLS habit system (Goodman et al., 2016). This slower, habit-based suppression stands in contrast to the fast, flexible control normally provided by the OFC.

4.4 Converging Roles Under Uncertainty

A key goal of this study was to assess how mPFC and OFC lesions affect responses to environmental uncertainty. By swapping the barrier length mid-session, we introduced a change that invalidated the stable physical context of the task. As expected, this manipulation robustly increased EFS in sham animals, confirming that this change successfully induced an information-seeking, exploratory response. The most striking finding, however, was the convergent effect of both mPFC and OFC lesions. Both groups exhibited a significantly blunted EFS response immediately following the swap compared to shams. This blunted response likely reflects a compound deficit. For the OFC, this aligns with its role in representing a "cognitive map of the task space," which allows for flexible behavior when environmental contingencies change (Wilson et al., 2014). A sudden change in the environment creates the kind of instability where the OFC is thought to be critical for updating outcome expectancies (Riceberg & Shapiro, 2012). Without a properly updated model of the task, the need to explore may not be registered. For the mPFC, the deficit is consistent with its established role in promoting behavioral flexibility, such as overriding a prepotent strategy when task rules change (Ragozzino, 2007) and specifically initiating the deliberate search for information (González et al., 2024). The failure of both lesion groups to mount a robust exploratory response suggests that both regions are critical for detecting environmental volatility and deploying an appropriate, adaptive strategy.

Furthermore, the data revealed a profound deficit in long-term adaptation in both lesion groups. While sham animals gradually reduced their EFS across the three sessions of the swap paradigm, both mPFC and OFC animals failed to do so. Relative to the adapting

sham animals, the lesioned animals maintained a persistently high and inflexible level of exploration across sessions. This failure to learn from recurring uncertainty is consistent with literature implicating both regions in long-term strategy development. Inactivation of the mPFC, for instance, impairs the ability to develop and maintain an optimal strategy in tasks with uncertain outcomes, like the Rat Gambling Task (Zeeb et al., 2015). The OFC is likewise essential for building and updating a model of the task structure (Wilson et al., 2014); a failure in this process would prevent the animal from learning that the mid-session swap is a predictable feature of the task, causing them to react to it as novel each day. This persistent reaction may reflect a failure to resolve a state of high "unexpected uncertainty," a computational concept describing surprising environmental changes that require a shift in strategy and has been strongly linked to prefrontal function (Yu & Dayan, 2005).

Taken together, these results reveal a convergent functional role for the mPFC and OFC under conditions of uncertainty, which stands in contrast to their dissociable roles in the stable version of the task. In a predictable environment, the OFC appears specialized for suppressing valueless, off-task exploration. However, when the task structure becomes unstable, both the mPFC and OFC seem to be recruited as part of a broader prefrontal network governing behavioral flexibility. This network appears responsible for both the immediate recognition that the environment has changed (requiring more exploration) and the long-term consolidation of a new adaptive strategy. The blunted immediate response and the failure to adapt across sessions in our lesioned animals suggest that without these key prefrontal nodes, rats are less responsive to change and are unable to learn from environmental instability.

4.5 Limitations and Future Directions

Several limitations of the present study should be acknowledged. The final sample sizes were modest, a consequence of subject attrition from both failures to meet training criteria and lesions that were incomplete or extended into adjacent structures. While our statistical models accounted for repeated measures to maximize power, future studies with larger group sizes would be beneficial to confirm these findings and detect more subtle effects. Beyond statistical power, a key conceptual limitation arises from the use of permanent excitotoxic lesions. While effective for establishing the necessity of a brain region for a given function, this method introduces the potential confound of functional remapping during the 10-day recovery period (Vaidya et al., 2019). The behaviors observed in this study therefore reflect the function of a reorganized system, not the acute, moment-to-moment contributions of these regions in an intact brain.

Future studies employing techniques with greater temporal and spatial precision are therefore essential to build upon these foundational findings. To circumvent the issue of long-term compensation, temporary inactivation using optogenetic or chemogenetic tools would allow for a more precise dissection of function. For instance, inactivating the mPFC or OFC only during the barrier-swap portion of the task could clarify their specific roles in responding to uncertainty without the confound of neural plasticity. Furthermore, moving beyond broad regional manipulations is critical. Given the functional heterogeneity within both the mPFC and OFC, future work should target specific subregions (e.g., PL vs. IL) or projection-specific pathways (e.g., OFC-striatal or mPFC-striatal circuits) to understand how distinct networks contribute to these complex behaviors.

Finally, in vivo electrophysiological recordings would be a powerful complement to the present lesion data. Monitoring mPFC and OFC ensembles as rats perform the task could reveal the moment-to-moment computations that give rise to the behavioral patterns uncovered here. Such recordings would allow us to study intact circuits involved in exploration, adding mechanistic detail to what lesion studies have shown already. Together, converging evidence would offer a more complete picture of how prefrontal networks govern adaptive decision-making under both stable and volatile conditions.

Chapter 5: Conclusion

In summary, this thesis provides evidence for distinct, dissociable roles of the medial prefrontal cortex (mPFC) and orbitofrontal cortex (OFC) in the regulation of exploratory behavior within a stable, well-learned environment. We found that the OFC is critical for suppressing inefficient, off-task exploration. Lesions to this region appeared to disconnect the animal's behavior from the learned zero-value of extraneous feeder sampling (EFS), resulting in persistently high initial exploration. In contrast, the mPFC appears to be involved in the strategic, moment-to-moment regulation of exploration. The accelerated decay of EFS within a session in mPFC-lesioned animals suggests an impairment in the ability to flexibly maintain an exploratory strategy, leading to its premature abandonment. Together, these findings delineate a division of labor in a stable environment: the OFC provides a value-based signal to gate exploration, while the mPFC dynamically governs the execution of the exploratory strategy over time.

However, this functional dissociation gave way to a striking convergence when the environment became uncertain. In response to a sudden change in the task's physical

configuration, both mPFC and OFC lesions produced a common set of deficits: a blunted immediate exploratory response and a profound failure to adapt to the recurring uncertainty across multiple sessions. This suggests that while these regions have specialized functions in a predictable world, they are both recruited as essential nodes in a broader prefrontal network responsible for managing environmental volatility. The ultimate contribution of this work is to demonstrate this context-dependent neural architecture: a system that relies on specialized modules for efficiency in stable conditions but engages a more distributed, parallel network to guide flexible, adaptive behavior when faced with uncertainty.

References

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4), 407–419. https://doi.org/10.1016/S0028-3908(98)00033-1

Birrell, J. M., & Brown, V. J. (2000). Medial Frontal Cortex Mediates Perceptual Attentional Set Shifting in the Rat. *Journal of Neuroscience*, *20*(11), 4320–4324. https://doi.org/10.1523/JNEUROSCI.20-11-04320.2000

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Brooks, M. E., Kristensen, K., Benthem, K. J. van, Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Capuzzo, G., & Floresco, S. B. (2020). Prelimbic and Infralimbic Prefrontal Regulation of Active and Inhibitory Avoidance and Reward-Seeking. *Journal of Neuroscience*, *40*(24), 4773–4787. https://doi.org/10.1523/JNEUROSCI.0414-20.2020

Caracheo, B. F., Emberly, E., Hadizadeh, S., Hyman, J. M., & Seamans, J. K. (2013). Abrupt changes in the patterns and complexity of anterior cingulate cortex activity

when food is introduced into an environment. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00074

Corbit, L. H., & Balleine, B. W. (2011). The General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer Are Differentially Mediated by the Nucleus Accumbens Core and Shell. *Journal of Neuroscience*, *31*(33), 11786–11794. https://doi.org/10.1523/JNEUROSCI.2711-11.2011

Costa, V. D., & Averbeck, B. B. (2020). Primate Orbitofrontal Cortex Codes Information Relevant for Managing Explore–Exploit Tradeoffs. *Journal of Neuroscience*, *40*(12), 2553–2561. https://doi.org/10.1523/JNEUROSCI.2355-19.2020

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. https://doi.org/10.1038/nature04766

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The Good, The Bad and The Ugly. *Current Opinion in Neurobiology*, *18*(2), 185–196. https://doi.org/10.1016/j.conb.2008.08.003

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The Role of Medial Prefrontal Cortex in Memory and Decision Making. *Neuron*, *76*(6), 1057–1070. https://doi.org/10.1016/j.neuron.2012.12.002

Featherstone, R. E., & McDonald, R. J. (2004). Dorsal striatum and stimulus-response learning: Lesions of the dorsolateral, but not dorsomedial, striatum impair

acquisition of a stimulus-response-based instrumental discrimination task, while sparing conditioned place preference learning. *Neuroscience*, *124*(1), 23–31. https://doi.org/10.1016/j.neuroscience.2003.10.038

Gabbott, P. L. A., Warner, T. A., Jays, P. R. L., Salway, P., & Busby, S. J. (2005). Prefrontal cortex in the rat: Projections to subcortical autonomic, motor, and limbic centers. *Journal of Comparative Neurology*, *492*(2), 145–177. https://doi.org/10.1002/cne.20738

González, V. V., Zhang, Y., Ashikyan, S. A., Rickard, A., Yassine, I., Romero-Sosa, J. L., Blaisdell, A. P., & Izquierdo, A. (2024). A special role for anterior cingulate cortex, but not orbitofrontal cortex or basolateral amygdala, in choices involving information. *Cerebral Cortex*, *34*(4), bhae135. https://doi.org/10.1093/cercor/bhae135

Goodman, J., Ressler, R. L., & Packard, M. G. (2016). The dorsolateral striatum selectively mediates extinction of habit memory. *Neurobiology of Learning and Memory*, *136*, 54–62. https://doi.org/10.1016/j.nlm.2016.09.012

Gourley, S. L., & Taylor, J. R. (2016). Going and stopping: Dichotomies in behavioral control by the prefrontal cortex. *Nature Neuroscience*, *19*(5), 656–664. https://doi.org/10.1038/nn.4275

Gruber, A. J., & Thapa, R. (2016). The Memory Trace Supporting Lose-Shift Responding Decays Rapidly after Reward Omission and Is Distinct from Other Learning Mechanisms in Rats. *eNeuro*, *3*(6). https://doi.org/10.1523/ENEURO.0167-16.2016

Gruber, A. J., Thapa, R., & Randolph, S. H. (2017). Feeder Approach between Trials Is Increased by Uncertainty and Affects Subsequent Choices. *eNeuro*, *4*(6). https://doi.org/10.1523/ENEURO.0437-17.2017

Hart, A. S., Rutledge, R. B., Glimcher, P. W., & Phillips, P. E. M. (2014). Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term. *Journal of Neuroscience*, *34*(3), 698–704. https://doi.org/10.1523/JNEUROSCI.2489-13.2014

Hartig, F. (2024). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. https://CRAN.R-project.org/package=DHARMa

Heidbreder, C. A., & Groenewegen, H. J. (2003). The medial prefrontal cortex in the rat: Evidence for a dorso-ventral distinction based upon functional and anatomical characteristics. *Neuroscience & Biobehavioral Reviews*, *27*(6), 555–579. https://doi.org/10.1016/j.neubiorev.2003.09.003

Hoover, W. B., & Vertes, R. P. (2007). Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Structure and Function*, *212*(2), 149–179. https://doi.org/10.1007/s00429-007-0150-4

Ivan, V. E., Banks, P. J., Goodfellow, K., & Gruber, A. J. (2018). Lose-Shift Responding in Humans Is Promoted by Increased Cognitive Load. *Frontiers in Integrative Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnint.2018.00009

Izquierdo, A. (2017). Functional Heterogeneity within Rat Orbitofrontal Cortex in Reward Learning and Decision Making. *The Journal of Neuroscience*, *37*(44), 10529–10540. https://doi.org/10.1523/JNEUROSCI.1678-17.2017

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, *4*, 237–285. https://doi.org/10.1613/jair.301

Killcross, S., & Coutureau, E. (2003). Coordination of Actions and Habits in the Medial Prefrontal Cortex of Rats. *Cerebral Cortex*, *13*(4), 400–408. https://doi.org/10.1093/cercor/13.4.400

Koralek, A. C., & Costa, R. (2021). Dichotomous dopaminergic and noradrenergic neural states mediate distinct aspects of exploitative behavioral states. *Science Advances*. https://doi.org/10.1126/SCIADV.ABH2059

Krebs, J. R., Kacelnik, A., & Taylor, P. (1978). Test of optimal sampling by foraging great tits. *Nature*, *275*(5675), 27–31. https://doi.org/10.1038/275027a0

Lee, D., Conroy, M. L., McGreevy, B. P., & Barraclough, D. J. (2004). Reinforcement learning and decision making in monkeys during a competitive game. *Cognitive Brain Research*, *22*(1), 45–58. https://doi.org/10.1016/j.cogbrainres.2004.07.007

Lee, D., Seo, H., & Jung, M. W. (2012). Neural Basis of Reinforcement Learning and Decision Making. *Annual Review of Neuroscience*, *35*(Volume 35, 2012), 287–308. https://doi.org/10.1146/annurev-neuro-062111-150512

McDonald, R. J., & White, N. M. (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience*, *107*(1), 3–22. https://doi.org/10.1037/0735-7044.107.1.3

McGillycuddy, M., Warton, D. I., Popovic, G., & Bolker, B. M. (2025). Parsimoniously Fitting Large Multivariate Random Effects in glmmTMB. *Journal of Statistical Software*, *112*(1), 1–19. https://doi.org/10.18637/jss.v112.i01

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, *24*(Volume 24, 2001), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947. https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996

Paxinos, G., & Watson, C. (2006). *The Rat Brain in Stereotaxic Coordinates: Hard Cover Edition*. Elsevier.

Ragozzino, M. E. (2007). The Contribution of the Medial Prefrontal Cortex, Orbitofrontal Cortex, and Dorsomedial Striatum to Behavioral Flexibility. *Annals of the New York Academy of Sciences*, *1121*(1), 355–375. https://doi.org/10.1196/annals.1401.013

Riceberg, J. S., & Shapiro, M. L. (2012). Reward Stability Determines the Contribution of Orbitofrontal Cortex to Adaptive Behavior. *Journal of Neuroscience*, *32*(46), 16402–16409. https://doi.org/10.1523/JNEUROSCI.0776-12.2012

Rich, E. L., & Shapiro, M. (2009). Rat Prefrontal Cortical Neurons Selectively Code Strategy Switches. *Journal of Neuroscience*, *29*(22), 7208–7219. https://doi.org/10.1523/JNEUROSCI.6068-08.2009

Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*(4), 389–397. https://doi.org/10.1038/nn2066

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neuroscience*, *1*(2), 155–159. https://doi.org/10.1038/407

Schoenbaum, G., Nugent, S. L., Saddoris, M. P., & Setlow, B. (2002). Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *NeuroReport*, *13*(6), 885.

Schreiner, D. C., & Gremel, C. M. (2018). Orbital Frontal Cortex Projections to Secondary Motor Cortex Mediate Exploitation of Learned Rules. *Scientific Reports*, *8*(1), 10979. https://doi.org/10.1038/s41598-018-29285-x

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Skelin, I., Hakstol, R., VanOyen, J., Mudiayi, D., Molina, L. A., Holec, V., Hong, N. S., Euston, D. R., McDonald, R. J., & Gruber, A. J. (2014). Lesions of dorsal striatum eliminate lose-switch responding but not mixed-response strategies in rats. *European Journal of Neuroscience*, *39*(10), 1655–1663. https://doi.org/10.1111/ejn.12518

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Tait, D. S., & Brown, V. J. (2007). Difficulty Overcoming Learned Non-reward during Reversal Learning in Rats with Ibotenic Acid Lesions of Orbital Prefrontal Cortex. *Annals of the New York Academy of Sciences*, *1121*(1), 407–420. https://doi.org/10.1196/annals.1401.010

Tervo, D. G. R., Kuleshova, E., Manakov, M., Proskurin, M., Karlsson, M., Lustig, A., Behnam, R., & Karpova, A. Y. (2021). The anterior cingulate cortex directs exploration of alternative strategies. *Neuron*, *109*(11), 1876-1887.e6. https://doi.org/10.1016/j.neuron.2021.03.028

Tervo, D. G. R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., & Karpova, A. Y. (2014). Behavioral Variability through Stochastic Choice and Its Gating by Anterior Cingulate Cortex. *Cell*, *159*(1), 21–32. https://doi.org/10.1016/j.cell.2014.08.037

Vaidya, A. R., Pujara, M. S., Petrides, M., Murray, E. A., & Fellows, L. K. (2019). Lesion studies in contemporary neuroscience. *Trends in Cognitive Sciences*, *23*(8), 653–671. https://doi.org/10.1016/j.tics.2019.05.009

Weiner, D. M., Levey, A. I., Sunahara, R. K., Niznik, H. B., O'Dowd, B. F., Seeman, P., & Brann, M. R. (1991). D1 and D2 dopamine receptor mRNA in rat brain. *Proceedings of the National Academy of Sciences*, *88*(5), 1859–1863. https://doi.org/10.1073/pnas.88.5.1859

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, *19*(1), 181–189. https://doi.org/10.1111/j.1460-9568.2004.03095.x

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience*, *22*(2), 513–523. https://doi.org/10.1111/j.1460-9568.2005.04218.x

Yu, A. J., & Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. *Neuron*, *46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Zeeb, F. D., Baarendse, P. J. J., Vanderschuren, L. J. M. J., & Winstanley, C. A. (2015). Inactivation of the prelimbic or infralimbic cortex impairs decision-making in the rat gambling task. *Psychopharmacology*, *232*(24), 4481–4491. https://doi.org/10.1007/s00213-015-4075-y