

**MODULATION OF PROBABILISTIC DISCOUNTING AND PROBABILISTIC  
REVERSAL LEARNING BY DOPAMINE WITHIN THE MEDIAL ORBITOFRONTAL  
CORTEX**

by

Nicole Lynn Jenni

B.Sc., The University of British Columbia, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

The Faculty of Graduate and Postdoctoral Studies  
(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA  
(Vancouver)

August 2017

© Nicole Lynn Jenni, 2017

## **Abstract**

Weighing the value of a reward against its likelihood of delivery is a necessary component of adaptive decision-making. The medial subregion of the orbitofrontal cortex (OFC) plays a key role in this form of cognition, as inactivation of this subregion in rats alters behaviour during risk/reward decision-making and a probabilistic assay of cognitive flexibility. The medial OFC receives dopaminergic input from midbrain neurons, yet whether dopamine (DA) modulates medial OFC function has been virtually unexplored. Here, we assessed how D<sub>1</sub> and D<sub>2</sub> receptors in the medial OFC may modulate adaptive decision-making in the face of probabilistic outcomes. One series of experiments assessed probabilistic reversal learning, while another set of studies assessed risk/reward decision-making using a probabilistic discounting task. Separate groups of well-trained rats, received intra-medial OFC microinfusions of selective D<sub>1</sub> or D<sub>2</sub> antagonists prior to task performance. Our results indicate that blocking D<sub>1</sub> receptors in the medial OFC impaired while blockade of D<sub>2</sub> receptors facilitated the number of reversals completed. This may be due to an impairment in probabilistic reinforcement learning, as effects were mediated by changes in errors during the initial discrimination of the task. One function for DA within the medial OFC might therefore be to inform about responses that yield a higher probability of reward over less profitable options to maintain adaptive choice. During risk/reward decision-making, blocking D<sub>1</sub> receptors reduced risky choice driven by an increase in negative feedback sensitivity. Blockade of D<sub>2</sub> receptors increased risky choice, mediated instead by an increase in reward sensitivity. This implicates medial OFC DA in dampening the win-stay/lose-shift strategy to limit the use of immediate reward feedback in situations where rats have prior knowledge about reward profitability. These findings highlight a novel role for medial OFC DA in guiding behavior in situations of reward uncertainty. Medial OFC D<sub>1</sub> and D<sub>2</sub> receptors play dissociable

and opposing roles in different forms of reward-related action selection. Elucidating how DA within different nodes of mesocorticolimbic circuitry biases behavior in these situations will expand our understanding of the mechanisms regulating optimal and aberrant decision-making.

## **Lay Summary**

Efficient decision-making involves weighing the costs and benefits associated with available options to maximize long-term profits. Activity in the medial orbitofrontal cortex, a subregion of the frontal lobes, is important for guiding adaptive decision-making, but considerably less is known about how dopamine, a neurochemical that gets released into this brain region, might act here to bias choice behaviour. In well trained rats, we blocked dopamine receptors in the medial orbitofrontal cortex prior to performance on a risk/reward decision-making task, or on an assay of behavioural flexibility. We found that dopamine in the medial orbitofrontal cortex can act at different receptors to bidirectionally control patterns of decision-making across both tasks. These studies confirm that dopamine signaling in the medial orbitofrontal cortex plays a central role in decision-making, and will serve as a first step in integrating this brain region within the broader neural circuitry that biases reward-seeking behaviour.

## **Preface**

All experiments were planned and carried out by Nicole Jenni at the University of British Columbia. Nicole Jenni completed all the surgeries. Nicole Jenni with the help of Yi Tao Li did the behavioural training, while Nicole Jenni performed all drug testing procedures. Nicole Jenni and Dr. Stan Floresco analyzed the data, while Nicole Jenni wrote this document with the editing help of Dr. Stan Floresco, the supervisory author on this project. Research for this thesis was approved by the Animal Care Committee at UBC, protocol number A14-0210.

## Table of Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Lay Summary</b> .....	<b>iv</b>
<b>Preface</b> .....	<b>v</b>
<b>Table of Contents</b> .....	<b>vi</b>
<b>List of Tables</b> .....	<b>viii</b>
<b>List of Figures</b> .....	<b>ix</b>
<b>Acknowledgements</b> .....	<b>x</b>
<b>Introduction</b> .....	<b>1</b>
Orbitofrontal cortex: anatomy and connectivity .....	1
Medial orbitofrontal cortex: behavioral functions .....	2
Medial orbitofrontal modulation: dopamine? .....	5
<b>Materials and Methods</b> .....	<b>7</b>
Animals .....	7
Apparatus .....	7
Lever pressing training .....	7
Probabilistic reversal learning task .....	9
Probabilistic discounting task .....	10
Reward magnitude discrimination .....	11
Surgery .....	12
Drugs and microinfusion procedure .....	12
Histology .....	13
Data analysis .....	14
<b>Results</b> .....	<b>18</b>
Dopaminergic modulation of probabilistic reversal learning within the medial OFC .....	18
Dopaminergic modulation of probabilistic discounting within the medial OFC .....	24
Reward magnitude discrimination .....	26
<b>Discussion</b> .....	<b>29</b>
Probabilistic reversal learning .....	30
Probabilistic discounting .....	33
Dopamine modulation of medial OFC function .....	36
Clinical implications .....	40

Summary and conclusions .....	41
<b>References.....</b>	<b>43</b>

## List of Tables

Table 1	Performance measures following $D_1$ and $D_2$ blockade during probabilistic reversal learning.....	24
Table 2	Performance measures following $D_1$ and $D_2$ blockade during probabilistic discounting, and the reward magnitude discrimination task .....	28

## List of Figures

Figure 1	Histology: acceptable placements residing within the medial OFC .....	14
Figure 2	Probabilistic reversal learning following D <sub>1</sub> and D <sub>2</sub> blockade .....	23
Figure 3	Probabilistic discounting and reward magnitude discrimination following D <sub>1</sub> and D <sub>2</sub> blockade.....	27

## **Acknowledgments**

I would like to extend a special thank-you to everyone who made this degree possible. First, I would like to thank my supervisor, Dr. Stan Floresco for your unwavering mentorship, and the many beers over which we kept these projects on track. Thank-you to my amazing lab mates, Maric, Gemma, Mieke, Einar, Patrick, Meagan, Courtney, Debra and Ryan for the support and friendships. A special thanks to my undergrads Olivia and Katie for their countless hours of work and a big thank-you to my family, friends and my GSS softball team for your support outside of the lab, I love each and every one of you. Finally, thanks to my committee members, Dr. Kiran Soma, and Dr. Rebecca Todd, your feedback and input is an invaluable part of my education, and I am truly appreciative of your time

## **Introduction**

Efficient decision-making involves weighing the costs and benefits associated with available outcomes in order to maximize long-term utility. It is considered a learned behavior, in which organisms integrate information regarding the profitability of previous choices to estimate expected values for future decisions (Rangel *et al*, 2008). These types of decisions are mediated in part by the orbitofrontal cortex (OFC) in both humans and animals. Pioneering studies into the neurobiology of decision-making were based primarily on experimental work conducted in human patients with brain lesions encompassing substantial portions of their OFC (Bechara *et al*, 1994; Bechara *et al*, 1996). Damage to this region resulted in a profile of behavior in which intellectual abilities, and executive functioning were well preserved, yet pathological abnormalities appeared in their ability to make advantageous decisions pertaining to their own lives (reviewed by Bechara, 2000). Following several decades of research attempting to ascertain the function of this brain region, it has more recently been proposed that clarifying the function of the OFC has been complicated by the fact that the lateral and medial subregions appear to be functionally dissociable – they often make different contributions to behavior by using mechanisms that are localized in anatomically separate divisions of this subregion of the prefrontal cortex (PFC) (Carmichael and Price, 1996; Noonan *et al*, 2012; Zald *et al*, 2014).

### **Orbitofrontal cortex: anatomy and connectivity**

The lateral and medial portions of the OFC show reliable differences in their connectivity. Carmichael and Price (1994;1995;1996) were among the first to map out the anatomical features of the OFC in non-human primates, showing that the OFC is made up of about 20 heterogeneous subregions that group themselves into two divisions, and that can be differentiated by the cortical projections they send and receive. The medial portions of the OFC belong to a more medial

network of prefrontal regions, while the more lateral regions share cortico-cortical projections with more dorsolateral regions. What's more, these dissociable prefrontal networks share very minimal connections with one another, and can further be differentiated by their subcortical connectivity. The lateral network is more heavily innervated by sensory modalities, while the medial network is more strongly interconnected with limbic structures such as the amygdala (Carmichael and Price, 1994; Carmichael and Price, 1995; Carmichael and Price, 1996).

The majority of this anatomical work has been done in nonhuman primates, and although similar mapping of cortico-cortical connectivity has not been as extensively studied in rodents, there is evidence that the patterns of orbitofrontal-subcortical connectivity are well preserved in rats. A similar medial/lateral principle governs the architecture of rodent OFC (Hoover and Vertes, 2011), and the subcortical projection patterns of the rat lateral vs medial OFC maintains a high degree of homology across species with the specific nuclei they innervate with targets like the amygdala, the striatum, and the hippocampus (Price, 2007; Ongür and Price, 2000; Heilbronner *et al*, 2016; Hoover and Vertes, 2011). With this in mind, it is clear that experiments that do not take these boundaries into account become much more difficult to interpret (Price, 2007).

### **Medial orbitofrontal cortex: behavioral functions**

On a functional level, studies that have taken the medial/lateral subdivisions into account, have provided fairly conclusive evidence that functional specializations do exist within the OFC, with the medial and lateral subregions making dissociable contributions to cognitive functions. Nevertheless, compared to the lateral portion, there have only been a handful of studies in animals that have tried to elucidate the specific contribution of the medial subregion of the OFC to cognition and behavior— even despite the surfacing body of literature that suggests the medial

subregion may be more important for guiding reward-related behavior, while the lateral OFC might play little to no role in the expression of goal-directed action (Balleine *et al*, 2011; Noonan *et al*, 2012; Ostlund and Balleine, 2007a; Ostlund and Balleine, 2007b; Stopper *et al*, 2014)

An essential component of goal-directed behavior is to consider both costs, and motivational drives (ie. reward magnitude, hunger etc.) for a reinforcer to assess its value. This value signal is encoded as a representation which can then be retrieved to direct response strategies that are in line with those assessments (Gourley *et al*, 2016; Ostlund and Balleine, 2007b). The medial OFC may play an important role in guiding goal-directed behaviors by retrieving value representations that are in line with expected outcomes. A consistent finding across behavioral studies is that lesions or inactivations of the medial OFC results in the adoption of choice-strategies based more on observable, or immediate reward feedback, rather than on internal value representations they should consult having had past experience with the available options (Bradfield *et al*, 2015; Dalton *et al*, 2016; Gourley *et al*, 2016; Mar *et al*, 2011; Stopper *et al*, 2014). In other words, the medial OFC might mediate the retrieval of outcome representations, when that information is necessary for choice between different goal-directed actions, and in specific situations where choice behavior should be biased based on previous learned experience, and is not directly observable. For example, lesioning the medial OFC creates deficits in outcome devaluation and Pavlovian-instrumental transfer effects when conducted in extinction (such that the outcome of an available option is no longer observable) requiring reliance on an internal action-outcome representation to make the best value-guided choice (Bradfield *et al*, 2015). Interestingly, medial OFC lesions did not affect performance on variations of these two tasks not conducted in extinction, such that outcomes can be directly observed. This means rats could use a strategy where they held outcome information in their

working memory, and did not need to access a stored representation. Based off these findings, it appears that intact medial OFC processing is necessary for updating and retrieving previously encoded outcome representations, but does not appear to be necessary for working memory, or the ability to hold in mind newly learnt (or immediately observable) outcomes (Bradfield *et al*, 2015). There is also evidence that the medial OFC uses these internal representations to guide appropriate effort expenditure relative to outcome value, as disrupting activity in this brain region increases break point ratios on a progressive ratio schedule of reinforcement for food (Gourley *et al*, 2010; Gourley *et al*, 2016).

The role of the medial OFC in goal-directed action seems to be particularly necessary in situations of reward uncertainty, when reward contingencies are probabilistic, and these value representations are likely unstable (Dalton *et al*, 2016; Stopper *et al*, 2014; Hall-McMaster *et al*, 2016; Clark *et al*, 2008). Human patients with damage to their medial OFC show riskier behavior on the Iowa Gambling Task (Bechara *et al*, 1994) and the Cambridge Gambling Task (Clark *et al*, 2008). While studies in rats have paralleled these findings, showing that inactivating the medial OFC during a probabilistic discounting task, in which rats choose between small/certain, and large/uncertain rewards, increases choice of the larger, probabilistic option. This effect was driven by an increase in win-stay behavior, wherein rats were more likely follow a risky win with another choice of the larger, probabilistic option (Stopper *et al*, 2014). This once again reflects the adoption of a strategy based on the immediate and observable reward feedback, rather than what these rats know regarding the profitability of the risky option at certain times. Inactivation of the medial OFC also impairs probabilistic reversal learning, driven in part by an increase in errors committed during the initial discrimination of the task (Dalton *et al*, 2016) highlighting that the medial OFC is necessary for the discrimination of probabilistic action-

outcome contingencies to guide goal-directed choice. Interestingly, intact medial OFC processing is not required for reversal learning with assured outcomes (Dalton *et al*, 2016), likely because choices can be made based on immediately observable outcomes, and would be less dependent on the retrieval of an internal representation of the “correct” option at that time. In this manner, it appears that it is the aspect of reward uncertainty or volatility in these situations that hijacks the recruitment of this brain region.

### **Medial orbitofrontal modulation: dopamine?**

The behavioral data to date clearly shows that the medial OFC mediates a complex set of cognitive mechanisms relating to value based decision-making in situations of uncertainty. Despite this mounting body of literature, we still know very little about the neurochemical modulation of medial OFC function. It is well established that DA transmission within other terminal nodes of the mesocortical DA system makes critical and dissociable contributions to reward related, and goal-directed behavior. For example, the probabilistic discounting task, which we now know is also sensitive to medial OFC disruption (Stopper *et al*, 2014), is primarily modulated by DA transmission. Dynamic changes in prefrontal DA tracks the amount of food reward obtained over time (St Onge *et al*, 2012). In this manner, prefrontal DA may signal changes in the frequency of rewarded actions or the relative utility of different options, and appears do so by control over dissociable networks of prefrontal neurons that are modulated by either the D<sub>1</sub> or D<sub>2</sub> receptor subtype, and that can be further dissociated by the output targets to which they signal. In this manner, dopaminergic tone on prefrontal D<sub>1</sub> receptors serves to reinforce actions yielding larger rewards via a network that interfaces with the nucleus accumbens, while DA acting on prefrontal D<sub>2</sub> receptors facilitates flexibility in decision biases via actions on neural networks that interface with the basolateral amygdala (Jenni *et al*, 2017).

D<sub>1</sub> and D<sub>2</sub> receptors in the frontal lobes also regulate many other cognitive functions like working memory and cognitive flexibility (Floresco *et al*, 2006; Seamans *et al*, 1998). In a similar vein, we know that disorders characterized by abnormal mesocortical DA, such as schizophrenia and Parkinson's disease show reliable deficits across several domains of cognitive functioning including goal directed, and reward related functions

With respect to mesocortical DA, the vast majority of this work in rats has focused on the medial PFC regions, yet the medial OFC also receives dopaminergic input from the ventral tegmental area (Oades and Halliday, 1987) and shares reciprocal connections with key nodes within the DA decision circuitry including the basolateral amygdala, the ventral striatum and other regions of the PFC (Carmichael and Price, 1996). In light of this, it is surprising that there are virtually no studies investigating how dopaminergic modulation within the medial OFC may bias cost-benefit analyses and other aspects of reward-related behaviour (Cosme *et al*, 2016). The aim in these sets of studies was to characterize a potential role for DA D<sub>1</sub>- and D<sub>2</sub>-like receptors in the medial OFC specifically during two behaviors critically dependent on both medial OFC function and DA transmission: probabilistic discounting to assess cost/benefit decision-making in situations of reward uncertainty; and probabilistic reversal learning to assess cognitive flexibility.

## **Materials and Methods**

### **Animals**

Adult male Long-Evans rats (Charles River Laboratories) weighing 225-300g at the start of the experiment were initially group housed and provided access to water and food *ad libitum* upon arrival while they familiarized themselves with the colony. They were handled daily for one week, and then food restricted to 85-90% of their free feeding weight prior to the onset of operant training. Their weights were monitored daily, and each rat was fed 14-18g of food at the end of each experimental day. Individual food intake was adjusted to maintain a steady but modest weight gain. The colony was maintained on a 12 hr light/dark cycle, with lights on at 7:00 a.m. The rats underwent behavioral testing between 8 a.m. and 12 p.m each day. All experiments were conducted in accordance with the Canadian Council on Animal Care guidelines regarding appropriate and ethical treatment of animals, and were approved by the Animal Care Centre at the University of British Columbia.

### **Apparatus**

Behavioral testing was conducted in operant chambers (31 x 24 x 21 cm; Med Associates) enclosed in sound-attenuating boxes. The chambers were equipped with a fan that provided ventilation and masked extraneous noise. A single 100 mA house light illuminated the chambers, and each chamber was fitted with two retractable levers located on each side of a central food receptacle in which 45 mg sweetened food reward pellets (Bioserv) were delivered by a dispenser. All data were recorded by a computer connected to the chambers via an interface.

### **Lever pressing training**

The initial training protocols described below were identical to those described in our previous

studies (St. Onge *et al*, 2012; Larkin *et al*, 2016). Rats were food restricted for a minimum of 3 days before starting operant training. The day before exposure to the operant chamber, each rat was given approximately 25 sugar reward pellets in their home cage to familiarize them with the reward. On the first day of training, two pellets were delivered into the food cup and crushed pellets were sprinkled on an extended lever before the rat was placed into the chamber. On consecutive days, rats were trained under a fixed-ratio 1 schedule to a criterion of 50 presses in 30 min on one lever and then again for the other side (counterbalanced). They then progressed to a simplified version of the full task. These 90 trial sessions began with the levers retracted and the operant chamber in darkness. A trial was initiated with the illumination of the house light and the insertion of one of the two levers into the chamber (randomized in pairs). Failure to respond on the lever within 10 s caused its retraction, the chamber to darken, and the trial was scored as an omission. A response within 10 s caused the lever to retract and the delivery of a single pellet with 50% probability. Rats who were to be trained on the probabilistic reversal learning task trained on a variant of this simplified version with an intertrial interval (ITI) of 20 seconds, whereas rats who were to be trained on the probabilistic discounting task trained on a variant with an ITI of 40 seconds. All rats were trained for approximately 3-4 d to a criterion of 80 or more successful trials (<10 omissions). Rats assigned to the probabilistic discounting experiments were tested for their side bias towards a particular lever immediately following the final session of retractable lever training. Past studies from our laboratory have shown that we could considerably reduce the number of training days by accounting for rat's innate side bias when designating the risky lever. This single session was made up of trials where both levers would be inserted into the chamber. On the first trial, a food pellet was delivered following a response made on either lever. Following this choice, food was delivered only after the rat

responded on the lever opposite to the one initially chosen. If the rat chose the same lever as the previous response, no food was delivered and the house light was extinguished. This would continue until the rat correctly chose the lever opposite to what he initially selected. After a response was made on each lever, a new trial would start, such that each trial in this side bias task consisted of at least one response on each lever. Rats received seven such trials, and would typically require 13-30 responses to complete the session. Their side bias was assigned based on the lever (left or right) selected most often during the initial choice of each trial. The only exception was if a rat happened to make a disproportionate number of responses on one lever over the entire session (ie. a 2:1 ratio for the total number of presses). In each case, the risky lever was assigned *opposite* to their determined side bias. On the following day, these rats started training on the probabilistic discounting task.

### **Probabilistic reversal learning task**

This task was modified from the procedures described by Bari et al, (2010), and are identical to those described in previous studies (Dalton *et al*, 2014; Dalton *et al*, 2016). This 50-minute task was made up of 200 discrete choice trials with a 15s ITI. Each trial began with illumination of the house light three seconds before both levers were inserted into the chamber. Prior to each daily session, one lever was randomly designated the “correct” lever, where choice of this option delivered a one sugar pellet reward on 80% of the trials. The other lever was designated the “incorrect lever” however, if the rats chose this option, they still had a 20% chance of receiving the one sugar pellet reward. Failure to respond within 10 seconds of lever insertion led to their retraction, and the termination of the house light. This was scored as an omission. Following eight consecutive choices of the “correct” lever, the animal scored their first reversal, and the reward contingencies would switch such that the “correct” lever became the “incorrect” lever

and vice versa. This pattern continued over the course of the 200 trials. Squads of rats were trained until they displayed stable choice behavior, determined by analyzing the number of reversals from three consecutive sessions with a one-way repeated-measures ANOVA. If there was no main effect of day (at  $p > 0.10$ ) then choice behavior of the group was deemed stable. Because this task requires considerably fewer training sessions to achieve stable choice behavior compared to the discounting task, these rats were implanted with guide cannulae prior to behavioral training, such that once attaining behavioral stability on the task, they progressed straight to the microinfusion drug protocol.

### **Probabilistic discounting task**

Each daily session consisted of 90 trials separated into five blocks of 18 trials, and took about 50 minutes to complete. Rats were trained 5-7 days per week. One lever was designated the large/risky lever, and the other the small/certain lever (based on their side bias following lever press training) and this designation remained consistent throughout training. The large/risky lever was assigned opposite to each rat's side bias. Each session began in darkness with both levers retracted. Trials began every 35 seconds with the illumination of the house light. Three seconds later one or both levers were inserted into the chamber. Each of the five blocks consisted of eight forced choice trials (where only one lever was presented, randomized in pairs), followed by 10 free-choice trials. If no response was made within 10 seconds of lever presentation, the levers retracted and the chamber reverted to the intertrial state (omission). Selection of a lever caused its immediate retraction. A choice of the small/certain lever delivered one pellet with 100% probability. Choice of the large/risky lever delivered a 4-pellet reward in a probabilistic manner that decreased systematically across blocks of trials, (100, 50, 25, 12.5, 6.25%). The actual probability of receiving the large reward was drawn from a set probability distribution, so

that on any given day, receipt of the large reward would vary. As such, rats may not have experienced the exact probability assigned to that block within each daily session; rather, the actual probabilities averaged across multiple training sessions would more closely approximated the set value.

Squads of rats were trained until they displayed stable choice behavior, determined by analyzing data from three consecutive sessions with a two-way repeated-measures ANOVA with day and trial block as factors. If there was no main effect of day and no day x block interaction (at  $p > 0.10$ ) then choice behavior of the group was deemed stable and they were placed back in free feeding conditions. Two or three days later, rats were subjected to surgery. Different squads of rats required between 17-24 days of training before displaying stable patterns of choice.

### **Reward magnitude discrimination**

As we have done previously, we determined *a priori* that if a manipulation reduced risky choice on the discounting task, we would test the effect of that same manipulation on a separate group of rats trained on a reward magnitude discrimination task. This task consisted of 48 trials partitioned into four blocks of two forced-choice and 10 free-choice trials (12 trials per block). The probabilistic nature of the task was removed, such that a choice on the large reward lever delivered four pellets, while a choice on the other lever would deliver one pellet, both with 100% probability. After 8-10 d of training, rats displayed a strong bias towards the larger reward and were subjected to microinfusion test days in the same fashion as the animals that tested on the probabilistic discounting task. Because this task requires considerably fewer training sessions to achieve stable choice behavior compared to the discounting task, these rats were surgically implanted with guide cannulae prior to behavioral training.

## **Surgery**

Rats were provided food *ad libitum* for a minimum of 1-3 d prior to surgery. Rats were given a subanesthetic dose of ketamine and xylazine (50 and 5 mg/kg respectively) and maintained on isoflurane for the duration of the procedure. Rats were stereotaxically implanted with bilateral 23-gauge stainless steel guide cannula targeted just above the medial OFC [anteroposterior = + 4.5 mm; medial-lateral =  $\pm$  0.7 mm from bregma; dorsoventral = -3.3 mm from dura]. Cannulae were held in place with four stainless steel screws and dental acrylic. Thirty gauge obturators were inserted into the guide cannula and remained in place until infusions were performed. The animals were given a minimum of one week to recover from surgery before behavioral training. Rats in the probabilistic discounting experiments were required to retrain for a minimum of five days and until their group reestablished stable patterns of choice behavior before progressing to the testing protocol.

## **Drugs and microinfusion procedure**

Once stable choice behavior was established, animals received a mock infusion prior to their next training session to familiarize them with the testing procedures. Obturators were removed, injectors were placed inside the guide cannula for two minutes, but no infusion was administered.

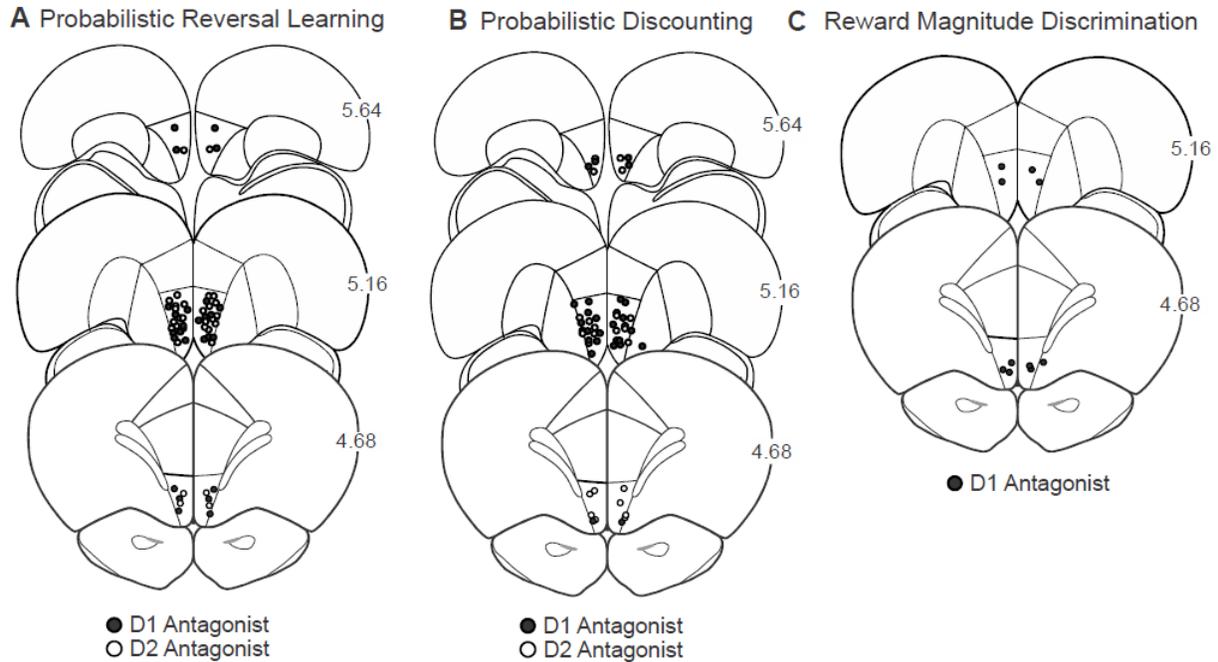
One or two days following the mock infusion, animals received their first microinfusion test day. Drugs or saline were infused at a volume of 0.5  $\mu$ l. The two DA drugs used in this study were: D<sub>1</sub> antagonist R-(+)-SCH23390 hydrochloride (0.1 and 1.0  $\mu$ g; Sigma-Aldrich) D<sub>2</sub> antagonist eticlopride hydrochloride (0.1 and 1.0  $\mu$ g; Sigma-Aldrich) dissolved in physiological (0.9%) saline. These doses were chosen based on previous studies that showed them to be

maximally effective at altering probabilistic discounting when infused bilaterally into other subregions of the PFC (St. Onge *et al*, 2011). Infusions were administered via 30-gauge injection cannulae that protruded 0.8mm past the end of the guide cannulae at a rate of 0.4  $\mu$ l/min by a microsyringe pump, so that the infusion lasted 75 s. The injection cannulae remained in place for one additional minute to allow for diffusion.

On the first test day, groups of rats received one of three infusions: a saline, a low and a high dose of their respective DA drug, 10 minutes prior to behavioral testing (a within-subjects design). The order of treatments was counterbalanced across animals. After the first infusion test day, animals were retrained for 1-3 days, making sure that their choice behavior deviated by <10% from their preinfusion baseline. They then received their second counterbalanced infusion, and this continued until each rat received all three treatments conditions.

### **Histology**

Rats were euthanized in a carbon dioxide chamber. Brains were fixed in a 4% formalin solution. Each brain was frozen and sliced in 50  $\mu$ m sections, mounted, and Nissl stained with Cresyl Violet. Placements were located with reference to the neuroanatomical atlas of Paxinos and Watson (2005) and can be seen in Figure 1 for all experiments. Data from rats who placements resided outside the borders of the medial OFC were removed from the analysis.



**Figure 1.** Histology. Schematic of coronal sections of the rat brain showing location of acceptable infusions in the medial OFC for rats in the **A**, Probabilistic reversal learning **B**, Probabilistic discounting and **C**, Reward magnitude discrimination experiments.

### Data analysis

*Probabilistic reversal learning:* The primary dependent variable of interest was the number of reversals completed per session, and was analyzed using a one-way repeated measures ANOVA. Secondary analyses were also conducted to clarify how a particular treatment affected reward and negative feedback sensitivity, the number of errors, and perseverative errors committed.

Reward and negative feedback sensitivity were indexed by win-stay and lose-shift ratios respectively. Win-stay ratios assessed the likelihood that a rat would follow a rewarded choice with another choice on the same lever regardless of whether this was a “correct” or “incorrect” choice. This was calculated from the number of trials on which a rat chose the same lever after being rewarded on the preceding trial divided by the total number of rewarded (correct or

incorrect) choices. Lose-shift ratios were based on the likelihood rats would shift to the other lever following omission of reward for their (correct or incorrect) choice on the previous trial. These ratios were analyzed using three-way repeated measures ANOVAs with treatment, response type (win-stay and lose-shift) and choice type (correct and incorrect) as three within-subjects factors.

The number of errors were tabulated for the initial discrimination (incorrect choices made before achieving the first criterion of 8 consecutive correct choices) and following the first reversal. This was analyzed with a two-way repeated measures ANOVA with treatment and phase (first discrimination and first reversal) as two within-subjects factors. Finally, we also analyzed the number of perseverative errors rats made during the reversal phases of the task. This was defined as the number of consecutive incorrect choices made following a reversal in reinforcement contingencies. As soon as the rat made one correct choice, subsequent errors were no longer considered perseverative in nature. We chose for this analysis to compare the number of perseverative errors made by each individual rat over the minimum number of reversals completed by that rat across the three treatments. We noticed that when rats complete a greater number of reversals, they tend to make fewer perseverative errors during the later reversal shifts, which had the potential to artificially reduce the number of perseverative errors for that treatment condition. For example, if a rat completed 6 reversals under control conditions, 5 reversals at a low dose of drug, and 3 reversals at the high dose of drug, we would compute the average number of perseverative errors made following the first 3 reversals for all treatments. These data were then subjected to a one-way repeated measure ANOVA. Latencies to make a choice and the number of trial omissions were also analyzed with one-way repeated measures ANOVAs. All

follow-up multiple comparisons were made using Dunnett's test where appropriate, because we were interested in whether any of the treatment conditions were different than control.

*Probabilistic discounting and reward magnitude discrimination:* The primary dependent variable of interest was the proportion of choices directed towards the large reward option, factoring out trial omissions. This was calculated in each block by dividing the number of choices of the large reward lever by the total number of trials in which the rats made a choice. Choice data were analyzed using a two-way within-subjects ANOVA with treatment and trial block as two within-subjects factors. In these analyses, the main effect of trial block was always significant ( $p < 0.001$ ), and simply indicative that rats did discount their choice of the high reward option as a function of large reward probability block, and will not be discussed further.

If a treatment induced a significant alteration in choice behavior on the probabilistic discounting task, we conducted a supplementary analysis to clarify whether these effects were attributable to changes in reward sensitivity (win-stay behavior) and/or negative-feedback sensitivity (lose-shift behavior). Each choice was analyzed according to the outcome of the preceding free-choice trial and expressed as a ratio. The win-stay score was calculated as a proportion of the number of risky choices made following a receipt of the larger reward (a risky win) divided by the total number of larger rewards obtained. Lose-shift scores were calculated as the proportion of small/certain choices made following a non-rewarded risky choice (risky loss) over the total number of non-rewarded risky choice trials. These scores were analyzed together using a two-way ANOVA, with response type (win-stay or lose-shift scores), and treatment as the two within-subject factors. Changes in win-stay behavior were used to index changes in reward sensitivity, and changes in lose-shift behavior served as an index of negative feedback sensitivity. In addition, response latencies, trial omissions, and locomotion were analyzed with

one-way repeated measures ANOVAs. Once again, all follow-up multiple comparisons were made using Dunnett's test where appropriate.

## Results

### Dopaminergic modulation of probabilistic reversal learning within the medial OFC

Previous work has shown that inactivation of the medial OFC impaired performance on the probabilistic reversal learning task (Dalton *et al*, 2016). The aim with this set of studies was to attempt to account for this effect by blocking dopaminergic modulation of the D<sub>1</sub> or the D<sub>2</sub> receptor subtype.

*D<sub>1</sub> receptor blockade:* One squad of rats trained on the probabilistic reversal learning task received counterbalanced infusions of a high or low dose of the D<sub>1</sub> antagonist (SCH 23390), or saline bilaterally into the medial OFC. Data from thirteen rats with acceptable cannulae placements residing within the medial OFC were included in the analysis (Fig 1A). Blockade of D<sub>1</sub> receptors impaired task performance, indexed by the number of reversals completed ( $F(2,24)=4.37, p=0.02$ , Fig 2A), for which follow up comparison using Dunnett's test confirmed a significant reduction in reversals for both the high and low dose treatments ( $p < 0.05$ ). To better understand the nature of this impairment, we assessed whether our manipulation influenced sensitivity to positive or negative feedback. We allotted each rat a win-stay score for the proportion of trials they maintained the same choice following receipt of reward (on both correct and incorrect trials) divided by their total number of wins to index their reward sensitivity. We also allotted each rat a lose-shift score calculated from the number of times a rat shifted choice following omission of reward divided by the total number of losses. This value served as a measure of negative feedback sensitivity, or how sensitive they were to omission of reward. We subjected these data to a three-way ANOVA with treatment, response type (win-stay, lose-shift) and choice type (correct, incorrect trials) as three within subjects factors. These analyses revealed that the impairment in reversals induced by D<sub>1</sub> blockade could not be attributed to any

changes in reward or negative feedback sensitivity. There was no significant main effect of treatment ( $F(2,24)=1.21, p=0.32$ ), or any interaction with the treatment factor (treatment x response type  $F(2,24)=2.01, p=0.16$ ; treatment x choice type  $F(2,24)=0.36, p=0.70$ , treatment x response type x choice type  $F(2,24)=0.16, p=0.85$ , Fig 2B). Even though D<sub>1</sub> receptor blockade within the medial OFC impaired probabilistic reversal performance, this did not appear to be driven by alterations in how the outcome of previous choices influenced subsequent action selection.

Additional analyses were designed to identify whether this reduction in reversals reflected an impairment during the reversal shift, or a more fundamental impairment in probabilistic discrimination learning. Specifically, we compared the number of errors committed by rats during the initial discrimination, as well as during the first reversal of the task. We analyzed these data using a two-way ANOVA with treatment and phase (initial discrimination, first reversal) as two within subjects factors. Analysis revealed a main effect of treatment ( $F(2,24)=4.97, p=0.02$ ), a main effect of phase. ( $F(1,12)=13.62, p=0.003$ ), but no treatment by phase interaction ( $F(2,24)=0.80, p=0.46$ ; Fig 2C). On average rats made more errors following the first reversal than during the initial discrimination, and more errors following infusion of the high (but not low) dose of SCH 23390 compared to their saline control (Dunnett's  $p<0.05$ ). We analyzed the number of perseverative errors, defined as the number of consecutive incorrect choices a rat committed following the switch in reinforcement contingencies after each completed reversal. There was no difference in the number of perseverative errors committed across treatment conditions ( $F(2,24)=2.05, p=0.15$ ; Fig 2D), suggesting an increase in perseveration cannot explain the impairment in reversals. Therefore, it appears that blockade of D<sub>1</sub> receptors in the medial OFC impairs the ability to use reward feedback to discriminate

between probabilistic schedules of reinforcement that are more vs less profitable. The lack of effect on perseveration suggests that the increase in errors seen following the first reversal is likely also due to an inability to identify and maintain choice on the now correct option, rather than an inability to disengage from the previous strategy. Lastly, blockade of D<sub>1</sub> receptors had no effect on trial omissions, locomotion, or response latencies (all  $F_s < 1.72$ , all  $p_s > 0.20$ ; Table 1).

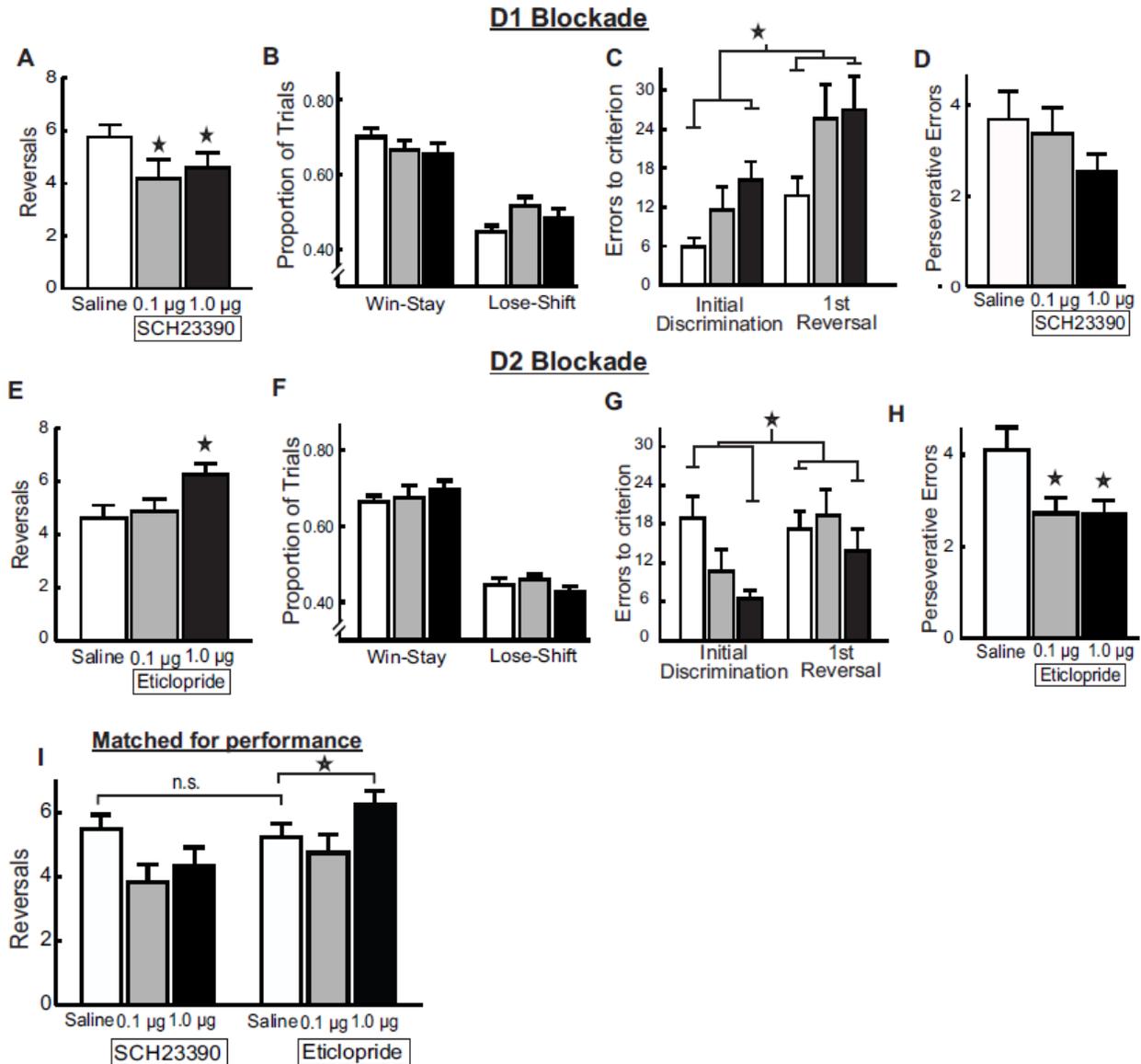
*D<sub>2</sub> receptor blockade:* Another 15 well-trained animals with acceptable placements (see Fig 1A) received infusions of the D<sub>2</sub> receptor antagonist eticlopride. In stark contrast to the effects of D<sub>1</sub> receptor blockade, D<sub>2</sub> antagonisms induced a seemingly opposite profile of behavior. This manipulation appeared to facilitate performance on this task, indexed by an increase in reversals ( $F(2,28)=5.38$ ,  $p=0.01$ ; Fig 2E), for which the high (but not low) dose of eticlopride caused an increase in reversals relative to the saline control (Dunnett's,  $p < 0.05$ ). Once again, this increase in reversals was not accompanied by any changes in reward or negative feedback sensitivity as our three-way ANOVA revealed no effect of treatment ( $F(2,28)=0.36$ ,  $p=0.70$ ) and no interactions with the treatment factor (treatment x response type  $F(2,28)=1.38$ ,  $p=0.27$ ; treatment x choice type  $F(2,28)=0.76$ ,  $p=0.48$ ; treatment x response type x choice type  $F(2,28)=0.15$ ,  $p=0.86$ ; Fig 2F).

Analysis of errors revealed a significant main effect of treatment ( $F(2,28)=4.25$ ,  $p=0.02$ ) but no main effect of phase ( $F(1,14)=3.89$ ,  $p=0.07$ ) and no treatment by phase interaction ( $F(2,28)=1.83$ ,  $p=0.18$ ; Fig 2G). Rats committed fewer errors after receiving the high (but not low) dose of eticlopride compared to the saline control (Dunnett's  $p < 0.05$ ). Visual inspection of the data shows this effect was primarily driven by a reduction in errors during the first discrimination. Interestingly, we found that this manipulation also facilitated performance by

reducing the number of perseverative errors ( $F(2,28)=3.47, p=0.04$ ; Fig 2H). Blockade of D<sub>2</sub> receptors had no effect on response latency, trial omissions, or locomotion (all  $F$ s  $<0.82$ , all  $p$ s  $>0.45$ ; Table 1).

When comparing the effects of SCH 23390 and eticlopride on probabilistic reversal performance, it is notable that rats in the D<sub>1</sub> antagonist group appeared to display better performance and completed a greater number of reversals following saline infusion compared to those in the D<sub>2</sub> antagonist group. Although this difference was not statistically significant ( $t_{(26)}=1.66, p=0.11$ ), we wanted to confirm that the opposing effects of D<sub>1</sub> vs D<sub>2</sub> antagonism were not artifacts attributable to differences in control levels of performance across the two groups. To do so, we analyzed data from subsets of animals in each of the two groups whose baseline performance were more comparable. In matching performance across the groups, the analysis removed data from three rats in the D<sub>2</sub> antagonist group that showed the poorest performance after saline infusions (who completed 1, 2, and 3 reversals), and removed data from one rat in the D<sub>1</sub> antagonist group that showed the best performance under saline (who completed 8 reversals). This gave us an equal number of animals in each group ( $n=12$ ) that displayed much more comparable performance after saline infusions (D<sub>1</sub> groups mean =  $5.50 \pm 0.42$ ; D<sub>2</sub> group mean =  $5.25 \pm 0.38$ ;  $t_{(22)}=0.42, p=0.68$ , see Fig 2I). When we analyzed the data from this subset of animals, the basic profile of changes after D<sub>1</sub> or D<sub>2</sub> antagonism in the medial OFC were still apparent. This analysis again revealed a significant increase in reversals completed following blockade of D<sub>2</sub> receptors ( $F(2,22)=3.02, p=0.03$ ), and a marginally significant reduction in reversals following blockade of D<sub>1</sub> receptors ( $F(2,22)=3.30, p=0.056$ ; see Fig 2I). From these findings, we conclude that the opposing effects on probabilistic reversal performance induced by

D<sub>1</sub> vs D<sub>2</sub> blockade in the medial OFC are unlikely to be attributable to differences in baseline performance across the two groups.



**Figure 2.** Blockade of D<sub>1</sub>-vs-D<sub>2</sub> receptors within medial OFC differentially impairs probabilistic reversal learning. **A**, Infusion of SCH23390 into the medial OFC (n=13) reduced the number of reversal completed. **B**, D<sub>1</sub> blockade did not influence win-stay or lose-shift behaviour following correct or incorrect choices. **C**, Errors to achieve criterion performance during the initial discrimination and first reversal phases. D<sub>1</sub> blockade increased errors made during the initial discrimination and following the first reversal. **D**, D<sub>1</sub> blockade did not affect perseverative errors throughout the task. **E**, Infusion of eticlopride into the medial OFC (n=15) increased the number of reversals completed. **F**, D<sub>2</sub> blockade did not influence win-stay or lose-shift behaviour following correct or incorrect choices. **G**, D<sub>2</sub> blockade reduced errors during the initial discrimination and first reversal phases. **H**, D<sub>2</sub> blockade reduced perseverative errors throughout the task. **I**, Data from a subset of rats whose baseline performance were more comparable (n=12 in both groups). The same profile of changes after D<sub>1</sub>/D<sub>2</sub> blockade are still apparent, suggesting the opposing effects on reversals are unlikely due to differences in baseline performance across the two groups. Error bars are SEM, asterisk denotes  $p < 0.05$  compared to saline.

Table 1	Mean (SEM)					
	Probabilistic Reversal Learning	Saline	Low Dose (0.1 µg)		High Dose (1 µg)	
<b>D<sub>1</sub> Antagonist - SCH23390</b>						
Response Latency (s)	0.79 (0.14)	0.73 (0.08)	0.73 (0.11)			
Trial Omissions	3.08 (1.54)	1.77 (0.78)	3.31 (1.92)			
Locomotion	1398 (168)	1421 (116)	1210 (175)			
<b>D<sub>2</sub> Antagonist - Eticlopride</b>						
Response Latency (s)	0.67 (0.08)	0.73 (0.09)	0.72 (0.07)			
Trial Omissions	4.20 (2.59)	2.40 (1.54)	5.33 (3.54)			
Locomotion	1634 (169)	1711 (124)	1659 (211)			

**Table 1.** Performance measures following D<sub>1</sub> and D<sub>2</sub> blockade during probabilistic reversal learning. Latencies are measured in seconds, and locomotion is indexed by photobeam breaks. Values displayed are mean (SEM). Asterisk denotes  $p < 0.05$  compared to saline.

### Dopaminergic modulation of probabilistic discounting within the medial OFC

Reversible inactivation of the medial OFC increased risky choice on a probabilistic discounting task, by increasing reward sensitivity, or the likelihood that rats would continue choosing risky following receipt of the larger, probabilistic reward (Stopper *et al*, 2014). This experiment sought to determine how D<sub>1</sub> or D<sub>2</sub> receptor activity within the medial OFC activity contributes to guiding this form of decision-making.

*D<sub>1</sub> receptor blockade:* Data from 14 rats with acceptable cannulae placements were included in the analysis (Fig 1B). In contrast to the effects of medial OFC inactivation, D<sub>1</sub> receptor blockade caused a reliable reduction in risky choice relative to saline. This was confirmed by a significant main effect of treatment ( $F(2,26)=4.56, p=0.02$ ) but no treatment x block interaction ( $F(8,104)=1.80, p=0.09$ ; Fig 3A). To further explore what was mediating the reduction in risky choice, we analyzed each choice as a function of the outcome of the previous trial. A two-way repeated measures ANOVA with treatment and response type (win-stay, lose-shift) as two within

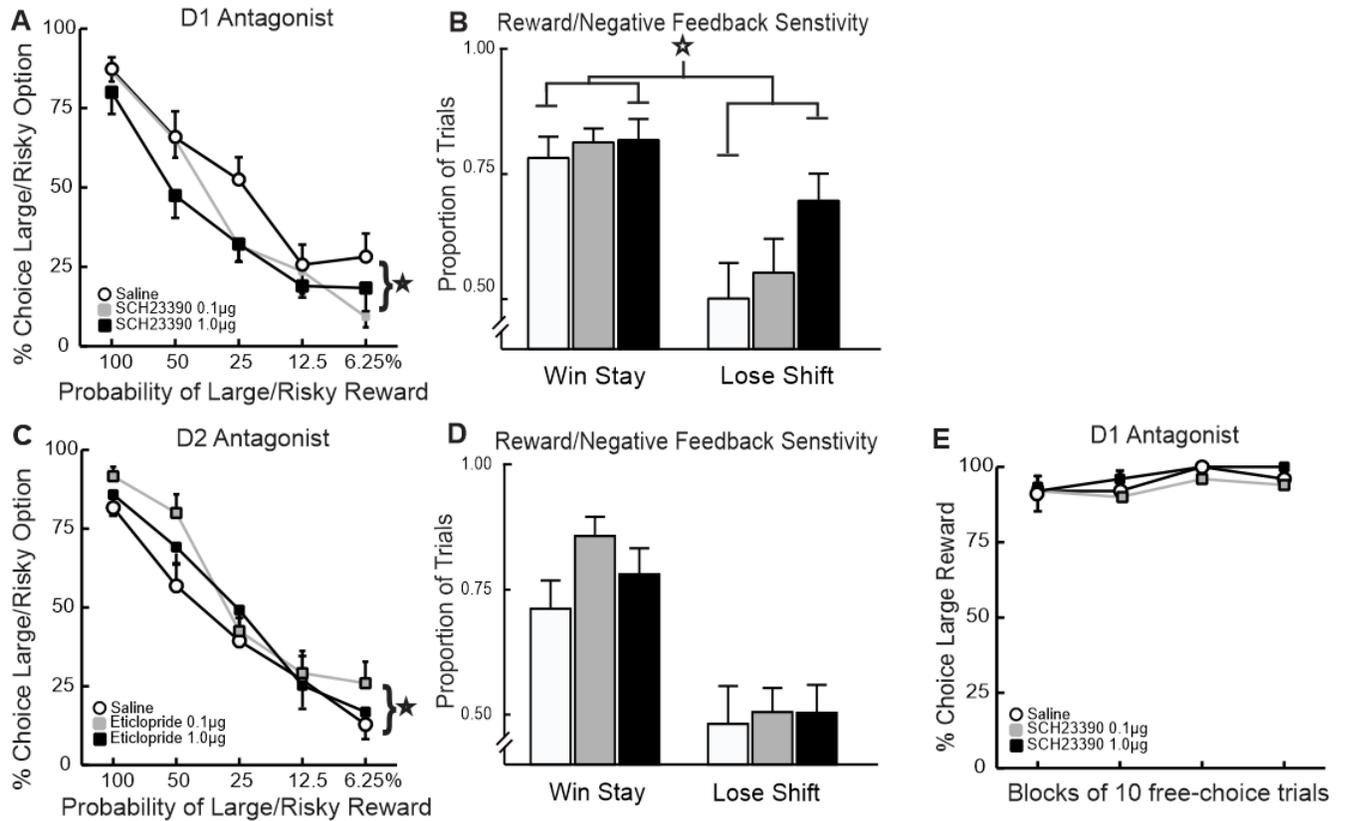
subjects factors revealed a significant main effect of treatment ( $F(2,24)=6.53, p=0.005$ ), but no interaction ( $F(2,24)=1.25, p=0.30$ ; Fig 3B). This main effect suggests that blocking  $D_1$  receptors in the medial OFC caused an overall increase in both win-stay/lose-shift behavior following the high (but not low) dose of SCH 23390 (Dunnett's  $p<0.05$ ). However, visual inspection of the data shown in Fig 3B indicates that the effects of these treatments on lose-shift behavior was much more prominent compared to the effects on win-stay behavior. This more pronounced increase in negative feedback sensitivity is likely to be the primary reason for the overall reduction in risky choice induced by  $D_1$  receptor blockade. In addition, blockade of  $D_1$  receptors also increased latency to make a choice following treatment with the high dose ( $F(2,26)=5.56, p=0.01$ ; Dunnett's  $p<0.05$ ) but did not affect locomotion ( $F(2,26)=0.84, p=0.44$ ) or trial omissions ( $F(2,26)=2.05, p=0.15$ ; see Table 2).

*D<sub>2</sub> receptor blockade:* Data from 12 rats with acceptable placements within the medial OFC were included in the analysis (Fig 1B). Disrupting  $D_2$  modulation of medial OFC activity was similar to the effects of medial OFC inactivation, but in contrast to blockade of  $D_1$  receptors in that these treatments caused an increase in risky choice relative to saline. The analysis of these data revealed a significant main effect of treatment ( $F(2,22)=3.82, p=0.04$ ; Fig 3C) but no treatment x block interaction ( $F(8,88)=0.87, p=0.55$ ). Curiously, multiple comparisons revealed that only the low dose induced a statistically significant increase risky choice relative to saline (Dunnett's  $p<0.05$ ). With respect to reward/negative feedback sensitivity,  $D_2$  receptor blockade caused an apparent increase in win-stay behavior, yet, the overall analysis of these data failed to yield a statistically significant main effect of treatment or interactions with the treatment factor (treatment main effect  $F(2,22)=1.59, p=0.23$ ; interaction  $F(2,22)=0.95, p=0.40$ ; Fig 3D). Despite this lack of an effect in the overall analysis, an exploratory comparison supported this

observation, revealing a significant increase in win-stay behavior in the low dose condition relative to the saline treatment ( $t_{(11)}=2.37$ ,  $p=0.04$ ), in a manner similar to what was observed following medial OFC inactivation (Stopper *et al*, 2014). There were no significant differences induced by blockade of D<sub>2</sub> receptors on locomotion, omissions, or decision latencies (all  $F_s < 2.11$ , all  $p_s > 0.15$ ; see Table 2).

### **Reward magnitude discrimination**

Disruption of D<sub>1</sub> modulation of medial OFC activity reduced preference for the larger probabilistic reward. In order to assess whether this was driven by a more general disruption in motivation, motor control, or the ability to discriminate between smaller vs larger rewards, we conducted a follow-up experiment with a separate cohort of rats trained on a simpler task. The reward magnitude discrimination task simply requires the rat to choose between two levers that deliver either a one or a four-pellet reward, both with 100% certainty. Five rats with acceptable placements in the medial OFC were trained for ~10 days after which they displayed a strong bias towards choosing the larger reward. The rats were then subjected to the same high and low dose infusions of the D<sub>1</sub> antagonist prior to task performance. Blockade of D<sub>1</sub> receptors did not alter choice (treatment main effect  $F(2,8)=0.19$ ,  $p=0.83$ ; interaction  $F(6,24)=0.13$ ,  $p=0.99$ ; Fig 3E) nor did it affect decision latencies, trial omissions or locomotion (all  $F_s < 0.92$ , all  $p_s > 0.43$ ; Table 2). The location of these infusions is displayed in Figure 1C. These data suggest that the effects from blocking D<sub>1</sub> receptors within the medial OFC are unlikely to be attributable to an impairment in discriminating between smaller and larger rewards, or other non-specific disruptions in motivational or motor processes.



**Figure 3.** Blockade of D<sub>1</sub> and D<sub>2</sub> receptors within the medial OFC differentially impairs probabilistic discounting. **A**, Percentage choice of the large/uncertain option following infusion of SCH23390 into the medial OFC across five blocks of free-choice trials. Blockade of D<sub>1</sub> receptors reduced choice for the larger reward. (n=14) **B**, Win-stay/lose-shift ratios. D<sub>1</sub> blockade increased sensitivity to losses. **C**, D<sub>2</sub> blockade in the medial OFC increased percentage choice of the large/uncertain option across five blocks of free-choice trials (n=12). **D**, Win-stay/lose-shift ratios. D<sub>2</sub> blockade caused no reliable differences across treatment conditions, although an exploratory direct comparison shows the increase in risky choice is likely driven by an increase in reward sensitivity at the low dose (0.1µg) of eticlopride. **E**, Disruption of D<sub>1</sub> modulation of medial OFC activity does not affect preference for larger versus smaller rewards on a simple reward magnitude discrimination. Error bars are SEM, asterisk denotes  $p < 0.05$  compared to saline.

<b>Probabilistic Discounting</b>	Mean (SEM)					
	<b>Saline</b>		<b>Low Dose (0.1 µg)</b>		<b>High Dose (1 µg)</b>	
<b>D<sub>1</sub> Antagonist - SCH23390</b>						
Response Latency (s)	0.71	(0.08)	0.67	(0.09)	0.92	(0.13)*
Trial Omissions	3.86	(1.89)	2.71	(1.28)	6.21	(2.56)
Locomotion	1698	(241)	1531	(251)	1453	(241)
<b>D<sub>2</sub> Antagonist - Eticlopride</b>						
Response Latency (s)	0.81	(0.14)	0.75	(0.11)	0.95	(0.21)
Trial Omissions	5.50	(2.74)	2.08	(1.13)	4.17	(1.97)
Locomotion	1409	(164)	1454	(150)	1374	(158)
<b>Reward Magnitude Discrimination</b>						
<b>D<sub>1</sub> Antagonist - SCH23390</b>						
Response Latency (s)	0.79	(0.11)	0.75	(0.10)	0.77	(0.11)
Trial Omissions	0	(0)	0	(0)	0	(0)
Locomotion	990	(112)	1081	(107)	1135	(92)

**Table 2.** Performance measures following D<sub>1</sub> and D<sub>2</sub> blockade during probabilistic discounting. Latencies are measured in seconds, and locomotion is indexed by photobeam breaks. Values displayed are mean (SEM). Asterisk denotes  $p < 0.05$  compared to saline

## Discussion

The goal with the present set of studies was to characterize a potential role for medial OFC D<sub>1</sub> and D<sub>2</sub> receptors during two behaviors known to be dependent on intact medial OFC processing: risk/reward decision making and a probabilistic assay of cognitive flexibility. Results showed that DA plays a causal role in modulating the medial OFC activity required for adaptive decision-making behavior, and that D<sub>1</sub> and D<sub>2</sub> receptors appear to hold a bidirectional control over behavior across multiple contexts. First, on the probabilistic reversal learning task, we found that blocking medial OFC D<sub>1</sub> receptors impaired, while blocking D<sub>2</sub> receptors facilitated the number of reversals made throughout the session. In both instances, these effects were in part driven by changes in the number of errors during the initial discrimination of the task – suggesting DA within the medial OFC might inform about responses that yield a higher probability of receiving reward over less profitable options in order to guide and maintain adaptive choices. Next, we found that blocking D<sub>1</sub> receptors reduced choice of the larger, probabilistic reward by increasing lose-shift behavior, suggesting that normal DA action on D<sub>1</sub> receptors promotes persistence in choice biases by dampening the impact that a reward omission has over subsequent choice. On the other hand, blocking D<sub>2</sub> receptors increased choice of the larger reward mediated instead by an increase in win-stay behavior – suggesting that D<sub>2</sub> receptors might instead function to mitigate the impact that receipt of the larger reward has over subsequent choice. In this manner, another potential function for DA within the medial OFC might be to limit the use of immediate reward feedback to guide choice in probabilistic situations where the animal has prior knowledge regarding the profitability at that time. Of note, DA within the medial OFC does not appear to mediate any fundamental changes in motivation, or motor control, nor does it impair the discrimination objectively larger rewards as we saw no effect of

D<sub>1</sub> blockade in deterministic settings, when the uncertainty costs on the larger reward were removed.

### **Probabilistic reversal learning**

This experiment provided novel insight into the dopaminergic modulation of medial OFC activity during probabilistic reversal learning, a task that assesses cognitive flexibility. D<sub>1</sub> vs D<sub>2</sub> receptor blockade induced dissociable and opposing profiles of behavior, with D<sub>1</sub> receptor blockade impairing probabilistic reinforcement learning, while D<sub>2</sub> receptor blockade appeared to facilitate this process.

*D<sub>1</sub> receptor blockade:* Blocking D<sub>1</sub> receptors impaired performance as indexed by a reduction in reversals completed. Although the number of reversals is typically used as the primary dependent variable to index behavioral flexibility, analysis of secondary measures supports the argument that this might not in fact be a profile of inflexibility *per se*. For one, we found this effect was driven in part by an increase in errors during the initial discrimination of the task, where rats need to use reward feedback to discriminate the “correct” from the “incorrect” option. Because reward contingencies were probabilistic, rats must track outcomes and hold in mind a representation of each option over multiple trials, to overcome any incongruent feedback (ie. a potential win on the incorrect lever) and maintain a bias towards the more profitable option. Blocking D<sub>1</sub> receptors impaired the use of probabilistic feedback to mediate this process, suggesting that DA acting at the D<sub>1</sub> receptor might inform the animal of responses that yield a higher probability of receiving reward over less profitable options in order to guide and maintain adaptive choices. This impairment in probabilistic discrimination following disruption to medial OFC activity is in line with findings in humans with damage to their medial OFC (Tsuchida *et al*,

2010) as well as in non-human primates with medial OFC lesions (Noonan *et al*, 2010), who reported probabilistic learning deficits in their subjects.

In addition, we found that D<sub>1</sub> blockade increased errors following the first reversal, however this was not accompanied by any changes in our measure of perseveration, suggesting that this might instead be driven by an increase in choice switching, or an impairment in maintaining an appropriate choice bias, rather than an impairment in disengaging from the previous action-outcome association. This is of interest because classic studies of OFC function, have typically associated this brain region with deficits in perseveration (Clarke *et al*, 2008; Gourley *et al*, 2010; Iversen and Mishkin, 1970; Rygula *et al*, 2010). This may in part be due to their defining boundaries of the OFC, as other nearby subregions such as the medial PFC for example have been associated with perseveration (Ragozzino, 2002), or it may be that their definition of perseveration is nonspecific. For example, Gourley *et al*, (2010) reported that lesioning the medial OFC caused an increase in responding for the incorrect option following a reversal in reward contingencies, and characterized this as an increase in perseveration for the previous strategy. The limited number of available options in this task however makes it hard to tell if this increase in errors following a reversal is due to the rats perseverating on the previous correct, but now incorrect strategy or if it instead might be due to a failure in identifying or maintaining the correct choice that is now more rewarding, as both phenotypes would lead to a higher number of errors following a reversal. This latter pattern of behavior is what Noonan and colleagues (2010; 2012) described in macaque monkeys with selective lesions of their medial OFC. Using a three-option probabilistic discrimination task, and a comprehensive analysis, they found that the failure to persevere with the new correct option following a reversal was a direct consequence of increased trial-by-trial switching as these animals were less likely to exploit a

strategy of repeating successful decisions. In line with these findings, Rudebeck and Murray (2008) also found when parsing apart errors following the first reversal, that the impairment seen following an OFC lesion was driven by errors committed after having made a new correct choice, once again arguing against a perseverative phenotype. In fact, medial OFC activity might instead be supporting the maintenance of correct choices (Clark *et al*, 2014), and the similarities in behavioral effects we found following D<sub>1</sub> receptor blockade suggests that DA might be serving as the modulating transmitter in this regard.

*D<sub>2</sub> receptor blockade:* Blocking D<sub>2</sub> receptors had a seemingly opposite effect on multiple performance measures associated with probabilistic reversal learning. It facilitated the number of reversals made throughout a session, with an effect that was once again driven in part by a reduction in initial discrimination errors. In this instance, blockade of D<sub>2</sub> receptors facilitated the use of probabilistic feedback to more quickly identify the more profitable option, and promoted the maintenance of a strategy for repeating successful decisions. The increase in reversals was accompanied by a reduction in perseverative errors, suggesting that this manipulation aided the rats to identify the change in contingencies in fewer trials. As a primarily inhibitory receptor, it may be that endogenous DA is acting at the D<sub>2</sub> receptor to dampen or suppress medial OFC activity underlying the maintenance and persistence of a choice bias, in favor of promoting a state of exploration of new options. The explore -vs- exploit tradeoff is a classic dilemma of uncertainty based decision-making (Wilson *et al*, 2014). Evolutionarily, it is conceivable that we should have a system in place to promote both exploration and exploitation in different situations, and it appears that DA within the medial OFC may be regulating this process, with D<sub>2</sub> blockade biasing this system towards an “exploitation-like” phenotype, whereby animals were more able to exploit the more profitable option in the first discrimination of the task.

Compared to a complete inactivation of the medial OFC (Dalton *et al*, 2016) this set of studies replicated the pattern of reversal and error effects with blockade at the D<sub>1</sub> receptor, but not the win-stay / lose-shift effects seen in that study. This leaves open the possibility that another transmitter system is informing medial OFC activity of the outcome of the previous trial to bias the next decision, whereas DA might be more involved in computing and comparing value to guide persistence or flexibility in choice. Indeed, it has been shown that altering serotonin transmission on a probabilistic reversal task quite similar to the one used here, changes patterns of reward and negative feedback sensitivity (Bari *et al*, 2010).

It is important to note that inactivating the medial OFC does not influence reversal learning during a similar task where reward contingencies were assured, highlighting that it is the probabilistic aspect of the task that recruits this cortical region (Dalton *et al*, 2016). Furthermore, we have seen previously, that the deficits induced by inactivation of the medial OFC are dissociable from other prefrontal structures, notably inactivation of the nearby prelimbic cortex of the medial PFC (Dalton *et al*, 2016) suggesting the changes in behavior we have seen in this study are unlikely due to drug diffusion into this nearby structure.

### **Probabilistic discounting**

*D<sub>1</sub> receptor blockade:* Perturbing D<sub>1</sub> modulation of medial OFC activity reduced choice of the large, uncertain reward, with this effect being driven by a substantial increase in lose-shift behavior. Following blockade of the D<sub>1</sub> receptor, rats became more sensitive to negative feedback, in that the reward omission following a risky loss had a more powerful control over subsequent behavior, increasing the probability they would shift towards the safer option on the next trial. Of note, this same manipulation during the deterministic reward magnitude discrimination task had no effect on choice behavior, suggesting that the effect of D<sub>1</sub> blockade

could not be attributed to a nonspecific motivational or spatial deficit, nor did it influence the ability for rats to correctly discriminate between smaller and larger rewards. These findings imply that normal tone on D<sub>1</sub> receptors within the medial OFC dampens negative feedback sensitivity to promote persistence in choice biases, by mitigating the impact that a reward omission has over subsequent choice. This pattern of behavior is consistent with the effects of blocking D<sub>1</sub> receptors in other terminal regions during this task – notably the prelimbic cortex of the medial PFC (St. Onge *et al*, 2011) and the nucleus accumbens (Stopper *et al*, 2013). It has recently been shown that these two regions form a functional PFC → nucleus accumbens circuit, that is under the modulatory control of prefrontal D<sub>1</sub> receptors (Jenni *et al*, 2017). The similarity of the medial OFC effects to the ones previously observed raises the intriguing possibility that the medial OFC may feed into this pathway, perhaps at the level of a cortico-cortical circuit. In this regard, it has been shown that DA receptors do modulate cortically projecting cells in cortical layer I (Wu and Hablitz, 2005; Happel, 2016), however, very little is known about DA anatomy in the medial OFC, so further research is required to better understand how this region is located within the broader circuitry regulating this behavior.

In contrast, blocking D<sub>2</sub> receptors in medial OFC increased choice of the risky option. This increase in risky choice was most apparent at the low dose of eticlopride, and was driven by an increase in preference for the larger reward in the 50% trial block, when there is the most uncertainty on whether the rat will win or lose. This effect, although not significant in the overall analysis, appeared to be driven by an increase in win-stay behavior. This was then confirmed by an exploratory direct comparison, suggesting receipt of the large reward increased the probability to continue choosing this option, or that that receipt of the larger reward had more motivational impact on subsequent choice. Interestingly, this profile of behavior is strikingly similar to what is

seen following a complete inactivation of the medial OFC (Stopper *et al*, 2014). It is worth noting that in that study, inactivating the medial OFC caused an increase in risky choice across two variants of the discounting task, one in which odds changed from 100 → 12.5% (as in the present study here) but also when odds increased from 12.5 → 100% (Stopper *et al*, 2014). This contrasts with inactivations of the prelimbic cortex, that induces differential patterns of choice. Specifically, medial PFC inactivation increases risky choice when odds change from good → bad, but reduces risky choice when odds increase from bad → good suggesting that the prelimbic region is important for adjusting decision biases when contingencies change (St. Onge and Floresco, 2010). These findings imply that normal DA tone on medial OFC D<sub>2</sub> receptors functions to dampen reward sensitivity, and mitigate the impact that a large reward has over subsequent choice behavior, and perhaps does not function to mediate flexibility, or the updating of decision biases when reward contingencies change, like D<sub>2</sub> receptors in the nearby medial PFC (St. Onge *et al*, 2011; Jenni *et al*, 2017). It is also important to highlight that previous studies have shown that inactivation of the lateral OFC does not affect choice behavior on this task, suggesting that this region too is unlikely mediating the observed effects (St. Onge and Floresco, 2010).

Overall, these findings highlight a novel role for medial OFC DA receptors in regulating cost/benefit decision making in situations of reward uncertainty. Furthermore, and of particular relevance to current theories of medial OFC function, this pattern of behavior suggests that DA within the medial OFC can act at different receptors to dampen the urge to make choices purely based on immediate reward feedback (a potential win or loss) from the previous trial, and instead encourages rats to consult previous knowledge on what they have learnt regarding the profitability of the large reward option at different times. In deterministic settings, a “win-stay

lose-shift” strategy where the rat treats each loss as indicative of a change in value, will maximize profitability; however, when reward feedback is determined by some probabilistic schedule, reliance on this strategy is highly maladaptive (Faraut *et al*, 2016). The DA signal in the medial OFC might serve to dampen this win-stay lose-shift strategy in situations of reward uncertainty, when a more advantageous strategy would instead be to consult an internal computation or representation built from previous experience with the available option. This is in line with the Bradfield *et al*, (2015) study on medial OFC function where they found that lesioning this brain region resulted in the adoption of choice strategies based on observable reward feedback, rather than on the consultation of an internal reward representation.

### **Dopamine modulation of medial OFC function**

Across both behaviors assayed in these sets of studies, DA signaling in the medial OFC plays a key role in guiding reward seeking in situations involving reward uncertainty. D<sub>1</sub> and D<sub>2</sub> receptors within the medial OFC appear to hold a dissociable and opposing influence in different forms of reward-related action selection. This feature of D<sub>1</sub> and D<sub>2</sub> receptors is certainly not novel. For one, D<sub>1</sub> family receptors are considered G<sub>s</sub>-coupled receptors, while D<sub>2</sub>-like receptors are G<sub>i</sub>-coupled receptors, that increase or decrease cAMP mediated signaling cascades (among others) respectively (Jackson and Westlind-Danielsson, 1994; Lachowicz and Sibley 1997). Activity at these receptors can also cause opposing actions on neuronal responses, for example a D<sub>1</sub> agonist can increase, while a D<sub>2</sub> agonist can decrease GABA release and NMDA currents in cortical neurons (Harsing and Zigmond, 1997; Starr, 1987; Zheng *et al*, 1999). Given this, it fits that medial OFC DA D<sub>1</sub> and D<sub>2</sub> receptors have an opposing influence on behavior across multiple contexts.

One particularly interesting feature of the present findings is that medial OFC D<sub>1</sub> receptor antagonism induced an effect similar to medial OFC inactivation on the probabilistic reversal task, whereas on the probabilistic discounting task, it was D<sub>2</sub> receptor blockade that produced an effect comparable to medial OFC inactivation. It is not clear why we replicated the two previous medial OFC inactivation studies (Dalton *et al*, 2016; Stopper *et al*, 2014) by blockade at different DA receptors. However, previous work has shown that in other prefrontal regions, there are dissociable networks of neurons that are under control of either the D<sub>1</sub> or D<sub>2</sub> receptor (Jenni *et al*, 2017), and that the resulting action of DA can have different effects depending on factors such as DA concentration, and the basal level of network activity (Seamans and Yang, 2004). If a similar principle governs the architecture of medial OFC neurons, it could be that dissociable networks of neurons are differentially modulated by D<sub>1</sub> or D<sub>2</sub> receptors and are recruited in distinct manners under the two task conditions assayed here. Although both tasks assess goal-directed behavior, there are fundamental differences in the types of information that are processed to guide action selection. The probabilistic discounting task requires a choice between rewards of differing magnitudes, one of which has an incurred uncertainty cost. On the other hand, the probabilistic reversal learning task requires a choice between different patterns of actions that may lead to rewards of a fixed magnitude. This could mean that there are different dopaminergic mechanisms that underlie the medial OFC mediation of choosing between rewards of different values, over choosing between different actions that may yield reward. This fundamental difference might also help explain the differences in win-stay and lose-shift effects seen across the two behaviors. It is interesting that in the probabilistic discounting task we found robust changes in reward and negative feedback sensitivity, yet we did not see the same patterns in the probabilistic reversal learning task. It seems, based on this difference, that medial OFC DA is

playing a greater role in modulating how recent action-outcomes modulate subsequent action selection when choosing between rewards of different subjective value, and not so much when choosing between different actions that may yield reward.

In trying to determine a unified theory on DA receptor function in the medial OFC, we can see there are similarities in the profiles we saw with our manipulations across both behaviors. In blocking D<sub>2</sub> receptors we saw an increased bias for the large reward option, mediated by an increase in reward sensitivity in the probabilistic discounting task, and a reduction in errors during the initial discrimination of the reversal learning task, suggesting they were faster to identify and maintain a strategy for the correct (or more profitable) option. Medial OFC D<sub>2</sub> blockade appears to support a more persistent, and stronger maintenance for decision biases in both instances. On the other hand, D<sub>1</sub> receptor blockade led to a reduction in bias for the large reward by increasing negative feedback sensitivity in the probabilistic discounting task, while it impaired reversal learning, driven by an increase in errors during the initial discrimination. Blocking D<sub>1</sub> receptors did not influence perseverative errors following the reversal shifts, suggesting that the impairment in reversal learning seen is not due to an inability to disengage from a previous rule, but instead an inability to maintain a choice bias for the now correct strategy. Across both behaviors, blockade of D<sub>1</sub> receptors appears to impair the identification and maintenance of a choice bias across subsequent trials.

Computational modeling of prefrontal DA modulation posits that mesocortical DA functions to transition neural networks between two activity states. One state is a highly persistent and stable activity state, with few changes in spontaneous activity, and is promoted by D<sub>1</sub> receptor stimulation. This occurs via D<sub>1</sub>-mediated changes in NMDA and GABA currents, making it harder for interfering stimuli or noise to interrupt activity (Seamans and Yang, 2004).

This persistent activity is thought to increase the robustness of working memory representations which underlies the D<sub>1</sub> receptor regulation of diverse functions. This state is thought to lock activity towards a single mode of action, such that one choice or outcome can dominate action selection even in the face of distractors, but this of course comes at the cost of response flexibility (Seamans and Yang, 2004). Blocking D<sub>1</sub> receptors, would by consequence bias this system towards less stable states, in which rats are less likely to identify and maintain patterns of behavior that lead to reward. This is in keeping with the finding that D<sub>1</sub> antagonism in the medial OFC impaired probabilistic learning during the initial discrimination of the reversal task, reduced risky choice and increased lose-shift behavior on the discounting task.

In contrast, stimulation of the D<sub>2</sub> receptor biases network activity away from robustness. It reduces NMDA and GABA currents (Zheng *et al*, 1999; Seamans *et al*, 2001), resulting in a network state where other activity tends to “pop” out spontaneously and can be easily disrupted (Durstewitz *et al*, 2000). This allows many items to be represented and compared simultaneously. This would likely support situations where it is advantageous to sample and compare different options, or in situations requiring response flexibility; but the downside of this state is that no option is particularly dominant in guiding behavior (Seamans and Yang, 2004). D<sub>2</sub> receptor blockade would therefore be expected to maintain patterns of behavior, which is in keeping with the enhanced probabilistic learning and increased choice of the larger reward observed here.

In this manner, DA acting preferentially on the D<sub>1</sub> or D<sub>2</sub> receptor subtype can bidirectionally alter activity states of prefrontal neural networks. Preferential receptor binding *in vivo* is biased in part on concentration dependent mechanisms, where lower concentrations act via D<sub>1</sub> receptors to enhance NMDA EPSCs and GABA IPSCs, but at higher concentrations, there

is greater D<sub>2</sub> receptor signaling that functions to reduce these currents (Zheng *et al*, 1999; Seamans *et al*, 2001; Trantham-Davidson *et al*, 2004). This likely suggests that dynamic fluctuations in medial OFC DA during these behaviors will bias whether there is preferential activity at the D<sub>1</sub> or D<sub>2</sub> receptor, and can ultimately bias actions selected by the organism. It is interesting to consider, that under this model, the DA signal might not be carrying information per se, as its main function is rather to transition the activity states of networks that can support various kinds of exploration -vs- exploitation behavior (Seamans and Yang, 2004). It is important to note that this model is based on how DA may modulate network states of the prelimbic region of the medial PFC. In comparison, there have been no studies probing how DA may regulate neural activity within the medial OFC. Nevertheless, given the similarities in the cellular and neurochemical make up of these two frontal lobe regions, it is plausible that the principles of operation of mesocortical DA modulation of network states that guide different patterns of behavior may be consistent across these two regions.

### **Clinical implications:**

Pathological changes in prefrontal DA transmission are at the root most cognitive deficits seen across psychiatric disorders such as schizophrenia and Parkinson's disease (Abi-Dargham *et al*, 2002; Robbins and Cools, 2014). The medial OFC has received particular attention as a site for cellular adaptations underlying maintenance and relapse in individuals with substance abuse disorders. For example the reduction in striatal D<sub>2</sub> receptors seen in methamphetamine addicted individuals is strongly related to changes in OFC metabolism (Volkow *et al*, 2001). While medial OFC gray matter is reduced in substance dependent individuals (Franklin *et al*, 2002) and this reduction persists years after abstinence (Tanabe *et al*, 2009). Additionally, and of particular therapeutic relevance, a recent study found blocking D<sub>1</sub> receptors in the rat medial OFC

completely abolished both cued- and cocaine primed-reinstatement of cocaine seeking behaviour (Cosme *et al*, 2016). These data suggest that restoring normal processing in the medial OFC holds potential as a therapeutic target for individuals suffering from substance abuse.

Characterizing and understanding how these understudied medial OFC circuits function in the normal brain is an important and necessary step in understanding how these circuits may drive pathological patterns of behavior seen in various psychiatric disorders, as there are currently very few studies in animals that have studied medial OFC function in this regard.

### **Summary and conclusions**

The findings reported here reveal a novel role for DA within the medial OFC in biasing goal-directed and reward related behavior in the face of probabilistic outcomes. We found that medial OFC D<sub>1</sub> and D<sub>2</sub> receptors play dissociable and opposing roles in biasing both cost/benefit analyses in situations of reward uncertainty, and action selection during a probabilistic assay of cognitive flexibility. Results from this set of studies adds to the argument that understanding OFC function will require isolating the dissociable and complimentary contributions of the medial and lateral portions to multiple aspects of behavior. There remain many questions in understanding the neurochemical mechanisms of medial OFC function, of primary interest is the output targets to which these DA receptors signal. It is possible that the medial OFC is regulating behavior via subcortical projections to targets like the nucleus accumbens and the basolateral amygdala, that it integrates via cortico-cortical projections with nearby regions like the medial PFC, or in more likelihood some combination of both. Nevertheless, these findings represent a first step in understanding how this brain region is located within the broader neural circuitry involved in biasing goal-directed action. Elucidating how DA within different nodes of

mesocorticolimbic circuitry biases behavior in these situations can expand our understanding of the mechanisms regulating both optimal and aberrant decision-making.

## References

- Abi-Dargham A, Mawlawi O, Lobardo I, Gil R, Martinez D, Huang Y *et al* (2002) Prefrontal dopamine D<sub>1</sub> receptors and working memory in schizophrenia. *The Journal of Neuroscience*, 22:3708-3719.
- Balleine BW, Leung BK, Ostlund SB (2011) The orbitofrontal cortex, predicted value, and choice. *Annals of the New York Academy of Sciences*, 1239:43–50.
- Bari A, Theobald DE, Caprioli D, Mar AC, Aidoo-Micha A, Dalley JW, *et al* (2010) Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology*, 35:1290–1301.
- Bechara A (2000) Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex*, 10:295–307.
- Bechara A, Damasio AR, Damasio H, Anderson SW (1994) Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50:7-15.
- Bechara A, Tranel D, Damasio H, Damasio AR (1996) Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex*, 6:215-225..
- Bradfield LA, Dezfouli A, van Holstein M, Chieng B, Balleine BW (2015) Medial orbitofrontal cortex mediates outcome retrieval in partially observable task situations. *Neuron*, 88:1268–1280.
- Carmichael ST, Price JL, (1994) Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *The Journal of Comparative Neurology*, 346:366-402.

- Carmichael ST, Price JL, (1995) Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *The Journal of Comparative Neurology*, 363:615-641.
- Carmichael ST, Price JL, (1996) Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *The Journal of Comparative Neurology*, 371:179–207.
- Clark L, Bechara A, Damasio H, Aitken MRF, Sahakian BJ, Robbins TW (2008) Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. *Brain*, 131:1311-1322.
- Clarke HF, Robbins TW, Roberts AC, (2008) Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *Journal of Neuroscience*, 28:10972–10982.
- Clarke HF, Cardinal RN, Rygula R, Hong TY, Fryer TD, Sawiak SJ *et al* (2014) Orbitofrontal dopamine depletion upregulates caudate dopamine and alters behaviour via changes in reinforcement sensitivity. *The Journal of Neuroscience*, 34:7663-7676.
- Cosme CV, Gutman AL, Worth WR, LaLumiere RT (2016) D<sub>1</sub> but not D<sub>2</sub> receptor blockade within the infralimbic and medial orbitofrontal cortex impairs cocaine seeking in a region-specific manner. *Addiction Biology*, Epub:1–12.
- Dalton GL, Phillips AG, Floresco SB, (2014) Preferential involvement by nucleus accumbens shell in mediating probabilistic learning and reversal shifts. *The Journal of neuroscience*, 34:4618–26.
- Dalton GL, Wang NY, Phillips AG, Floresco SB (2016) Multifaceted contributions by different regions of the orbitofrontal and medial prefrontal cortex to probabilistic reversal learning.

*The Journal of neuroscience*, 36:1996–2006.

Durstewitz D, Seamans JK, Sejnowski TJ, (2000) Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of Neurophysiology*, 83:1733–1750.

Faraut CM, Procyk E, Wilson CRE, (2016) Learning to learn about uncertain feedback. *Learning & Memory*, 23:90–99.

Floresco SB, Magyar O, Ghods-Sharifi S, Vexelman C, Tse MTL (2006) Multiple dopamine receptor subtypes in the medial prefrontal cortex of the rat regulate set-shifting. *Neuropsychopharmacology*, 31:297-309.

Franklin TR, Acton PD, Maldjian JA, Gray JD, Dackis CA, O'Brien CP *et al* (2002) Decreased gray matter concentration in the insular, orbitofrontal, cingulate and temporal cortices of cocaine patients. *Biological Psychiatry*, 51:134-142.

Gourley SL, Lee AS, Howell JL, Pittenger C, Taylor JR (2010) Dissociable regulation of instrumental action within mouse prefrontal cortex. *European Journal of Neuroscience*, 32:1726–1734.

Gourley SL, Zimmermann KS, Allen AG, Taylor JR, (2016) The medial orbitofrontal cortex regulates sensitivity to outcome value. *Journal of Neuroscience*, 36:4600–4613.

Hall-McMaster S, Millar J, Ruan M, Ward RD (2016) Medial orbitofrontal cortex modulates associative learning between environmental cues and reward probability. *Behavioral Neuroscience*, 131:1–10.

Happel MFK (2016) Dopaminergic impact on local and global cortical circuit processing during

- learning. *Behavioural Brain Research*, 299:32–41.
- Harsing LG, Zigmond MJ (1997) Influence of dopamine on GABA release in striatum: Evidence for D<sub>1</sub>-D<sub>2</sub> interactions and non-synaptic influences. *Neuroscience*, 77:419–429.
- Heilbronner SR, Rodriguez-Romaguera J, Quirk GJ, Groenewegen HJ, Haber SN (2016) Circuit-based corticostriatal homologies between rat and primate. *Biological Psychiatry*, 80:509–521.
- Hoover WB, Vertes RP (2011) Projections of the medial orbital and ventral orbital cortex in the rat. *Journal of Comparative Neurology*, 519:3766-3801.
- Iversen SD, Mishkin M (1970) Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Experimental Brain Research*, 11:376–386.
- Jackson DM, Westlind-Danielsson A (1994) Dopamine receptors: molecular biology, biochemistry and behavioural aspects. *Pharmacology & Therapeutics*, 64:291-370.
- Jenni NL, Larkin JD, Floresco SB (2017) Prefrontal D<sub>1</sub> and D<sub>2</sub> receptors regulate dissociable aspects of decision-making via distinct ventral striatal and amygdalar circuits. *Journal of Neuroscience*, [Epub ahead of print].
- Lachowicz JE, Sibley DR (1997) Molecular characteristics of mammalian dopamine receptors. *Pharmacology and Toxicology*, 81:105-113.
- Mar AC, Walker AL, Theobald DE, Eagle DM, Robbins TW (2011) Dissociable effects of lesions to orbitofrontal cortex subregions on impulsive choice in the rat. *Journal of Neuroscience* 31:6398–6404.
- Noonan MP, Walton ME, Behrens TE, Sallet J, Buckley MJ, Rushworth MF (2010) Separate

- value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 107:20547–20552.
- Noonan MP, Kolling N, Walton ME, Rushworth MF (2012) Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. *European Journal of Neuroscience*, 35:997–1010.
- Oades RD, Halliday GM (1987) Ventral tegmental (A10) system: neurobiology. 1. Anatomy and connectivity. *Brain Research Reviews*, 12:117–165.
- Ongür D, Price JL, (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral cortex*, 10:206–219.
- Ostlund SB, Balleine BW, (2007a) Orbitofrontal cortex mediates outcome encoding in pavlovian but not instrumental conditioning. *Journal of Neuroscience*, 27:4819–4825.
- Ostlund SB, Balleine BW, (2007b) The contribution of orbitofrontal cortex to action selection. *Annals of the New York Academy of Sciences*, 1121:174–192.
- Paxinos G, Watson C (2005) The rat brain in stereotaxic coordinates, Edition 5. San Diego Elsevier Academic.
- Price JL (2007) Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. *Annals of the New York Academy of Sciences*, 1121:54–71.
- Ragozzino ME (2002) The effects of dopamine D<sub>1</sub> receptor blockade in the prelimbic – infralimbic areas on behavioral flexibility. *Learning and Memory*, 9:18–28.

- Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9:545–556.
- Robbins TW, Cools R (2014) Cognitive deficits in parkinson's disease: a cognitive neuroscience perspective. *Movement Disorders* 29:597-607.
- Rudebeck PH, Murray EA (2008) Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *Journal of Neuroscience*, 28:8338–8343.
- Rygula R, Walker SC, Clarke HF, Robbins TW, Roberts AC (2010) Differential contributions of the primate ventrolateral prefrontal and orbitofrontal cortex to serial reversal learning. *Journal of Neuroscience*, 30:14552–14559.
- Seamans JK, Yang CR (2004) The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in Neurobiology*, 74:1–57.
- Seamans JK, Gorelova N, Durstewitz D, Yang CR (2001) Bidirectional dopamine modulation of GABAergic inhibition in prefrontal cortical pyramidal neurons. *The Journal of neuroscience*, 21:3628–3638.
- Seamans JK, Floresco SB, Phillips AG (1998) D<sub>1</sub> receptor modulation of hippocampal-prefrontal cortical circuits integrating spatial memory with executive functions in the rat. *The Journal of Neuroscience* 18:1613-1621
- St. Onge JR, Floresco SB (2010) Prefrontal cortical contribution to risk-based decision making. *Cerebral Cortex*, 20:1816–1828.
- St. Onge JR, Abhari H, Floresco SB (2011) Dissociable contributions by prefrontal D<sub>1</sub> and D<sub>2</sub> receptors to risk-based decision making. *Journal of Neuroscience*, 31:8625–8633.

- St. Onge JR, Ahn S, Phillips AG, Floresco SB (2012) Dynamic fluctuations in dopamine efflux in the prefrontal cortex and nucleus accumbens during risk-based decision making. *Journal of Neuroscience*, 32:16880–16891.
- Starr M (1987) Opposing roles of dopamine D<sub>1</sub> and D<sub>2</sub> receptors in nigral γ- [ 3h ] aminobutyric acid release? *Journal of Neurochemistry*, 49:1042–1049.
- Stopper CM, Khayambashi S, Floresco SB (2013) Receptor-specific modulation of risk-based decision making by nucleus accumbens dopamine. *Neuropsychopharmacology*, 38:715–728.
- Stopper CM, Green EB, Floresco SB (2014) Selective involvement by the medial orbitofrontal cortex in biasing risky, but not impulsive, choice. *Cerebral Cortex*, 24:154–162.
- Tanabe J, Tregellas JR, Dalwani M, Thompson L, Owens E, Crowley T, Banich M (2009) Medial orbitofrontal cortex gray matter is reduced in abstinent substance-dependent individuals. *Biological Psychiatry*, 65:160-164.
- Trantham-Davidson H, Neely LC, Lavin A, Seamans JK (2004) Mechanisms underlying differential D<sub>1</sub> versus D<sub>2</sub> dopamine receptor regulation of inhibition in prefrontal cortex. *The Journal of Neuroscience*, 24:10652–10659.
- Tsuchida A, Doll BB, Fellows LK (2010) Beyond reversal: A critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *Journal of Neuroscience*, 30:16868–16875.
- Volkow ND, Chang L, Wang GJ, Fowler JS, Ding YS, Sedler M *et al* (2001) Low level of brain dopamine D<sub>2</sub> receptors in methamphetamine abusers : association with metabolism in the

orbitofrontal cortex. *American Journal of Psychiatry*, 158:2015–2021.

Wilson RC, Geana A, White JM, Ludvig EA, Cohen JD (2014) Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, 143:2074–2081.

Wu J, Hablitz J (2005) Cooperative activation of D<sub>1</sub> and D<sub>2</sub> dopamine receptors enhances a hyperpolarization-activated inward current in layer I interneurons. *Journal of Neuroscience*, 25:6322–6328.

Zald DH, McHuge M, Ray KL, Glahn DC, Eickhoff SB, Laird AR (2014) Meta-analytic connectivity modeling reveals differential functional connectivity of the medial and lateral orbitofrontal cortex. *Cerebral Cortex*, 24:232–248.

Zheng P, Zhang XX, Bunney BS, Shi WX (1999) Opposite modulation of cortical N-methyl-D-aspartate receptor-mediated responses by low and high concentrations of dopamine. *Neuroscience*, 91:527–535.