Sex-dependent contributions of ventrolateral orbitofrontal cortex and basolateral amygdala to learning under uncertainty

Running Title: *stimulus and action-based reversal learning*

C.G. Aguirre[1]#, J.H. Woo[2]#, J.L.R Sosa[1], J. J. Munier[3], J. Perez[1], M. Goldfarb[1], K. Das[1], M. Gomez[1], T. Ye[1], J. Pannu[3], K. Evans[1], P.R. O'Neill[4], I. Spigelman[3], A. Soltani[2], and A. Izquierdo[1]

[1]Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095
[2]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755
[3]Section of Biosystems and Function, School of Dentistry, University of California, Los Angeles, Los Angeles, CA 90095
[4]Shirley and Stefan Hatos Center for Neuropharmacology, Department of Psychiatry and Biobehavioral Sciences, University of California Los Angeles, Los Angeles, CA 90095
#**co-first authors**

Author contributions: CGA and AI designed the research; CGA, JLRS, JM, JP, MG, KD, MG, TY, KE, and JP performed the research; CGA, JHW, JM, AS and AI analyzed the data; CGA, JHW, PRO, IS, AS, and AI interpreted the data; AI, AS, PRO, and IS acquired funding for the project; CGA and AI wrote the paper; all authors edited the final version.

Acknowledgements: This work was supported by UCLA's Division of Life Sciences Retention fund (Izquierdo), National Institutes of Health grants R01 DA047870 (Izquierdo and Soltani), R21 MH122800 (Izquierdo and Blair), R01AA024527 (Spigelman), K01 DA042219 (O'Neill), and F31 AA028183 (Munier), the NSF GRFP, Cota-Robles Fellowship, and Charles E. and Sue K. Young Fellowship (Aguirre), Ursula Mandel Fellowship and Graduate Research Mentorship award (Romero-Sosa), and the Training program in Neurotechnology Translation T32 NS115753 (Ye). We acknowledge the Staglin Center for Brain and Behavioral Health for additional support related to fluorescence microscopy. We also thank the NIDA Drug Supply program for the supply of clozapine-N-oxide.

Correspondence: *Alicia Izquierdo, Ph.D. email: aizquie@psych.ucla.edu Ph: +1 310 825 3459

Manuscript Information:

Pages 47
Figures 5 Extended Data Figures 5

Abstract

Reversal learning measures the ability to form flexible associations between choice outcomes with stimuli and actions that precede them. This type of learning is thought to rely on several cortical and subcortical areas, including highly interconnected orbitofrontal cortex (OFC) and basolateral amygdala (BLA), and is often impaired in various neuropsychiatric and substance use disorders. However, unique contributions of these regions to stimulus- and action-based reversal learning have not been systematically compared using a chemogenetic approach and particularly before and after the first reversal that introduces new uncertainty. Here, we examined the roles of ventrolateral OFC (vlOFC) and BLA during reversal learning. Male and female rats were prepared with inhibitory DREADDs targeting these regions and tested on a series of deterministic and probabilistic reversals during which they learned about stimulus identity or side (left or right) associated with different reward probabilities. Using a counterbalanced within-subject design, we inhibited these regions prior to reversal sessions. We measured initial and pre-post reversal changes in accuracy to measure first detection and further adjustment to reversals, respectively. We found that inhibition of vlOFC, but not BLA, eliminated detection of stimulus-based reversals. Conversely, both BLA and vlOFC inhibition resulted in significantly slower action-based reversal learning in females, not males, indicating a sex-dependent role for these regions in this type of learning. Further, learning in females was more impacted in first reversal by vlOFC inhibition than inhibition of BLA, the latter more involved in probabilistic reversal learning. These findings add to mounting evidence of sex-dependent learning flexibility.

Keywords: **stimulus learning, action learning, deterministic, probabilistic, reward learning, DREADDs**

63 **Significance Statement**

64 Inflexible learning is a feature of several neuropsychiatric disorders. We investigated how the
65 ventrolateral orbitofrontal cortex (vlOFC) and basolateral amygdala (BLA) are involved in
66 learning of stimuli or actions under different forms of uncertainty. Following chemogenetic
67 inhibition of these regions in both male and females, we measured detection and adjustment to
68 fully-predictive (i.e., deterministic) and subsequent probabilistic reversals. For action learning,
69 vlOFC and BLA exhibited a sex-dependent role in deterministic and probabilistic learning with
70 females exhibiting the slowest learning following inhibition. For stimulus learning, vlOFC, but not
71 BLA, was required for reversal detection and adjustment. These findings provide insight into the
72 mechanisms of learning under different forms of uncertainty and the sex-dependency of these
73 mechanisms.
74

Introduction

Reversal learning, impacted in various neuropsychiatric conditions, measures subjects' ability to form flexible associations between stimuli and actions with outcomes (Schoenbaum et al., 2003; Izquierdo et al., 2013; Dalton et al., 2016). Reversal learning tasks can also be used to probe learning following expected and unexpected uncertainty in the reward environment (Behrens et al., 2007; Jang et al., 2015; Winstanley and Floresco, 2016; Soltani and Izquierdo, 2019). For example, after the experience of the first reversal, all others are expected to some extent (Jang et al., 2015). Additionally, unexpected uncertainty can be introduced by changes in reward probabilities, after taking the baseline, expected uncertainty into account.

The basolateral amygdala (BLA) is an area of interest in reversal learning due its involvement in value updating (Tye and Janak, 2007; Janak and Tye, 2015; Wassum and Izquierdo, 2015; Groman et al., 2019) and the encoding of both stimulus-outcome and action-outcome associations typically probed in Pavlovian-to-Instrumental tasks (Corbit and Balleine, 2005; Lichtenberg et al., 2017; Malvaez et al., 2019; Sias et al., 2021). Manipulations of amygdala and specifically BLA have resulted in reversal learning impairments (Schoenbaum et al., 2003; Churchwell et al., 2009; Groman et al., 2019), impaired learning from positive feedback (Costa et al., 2016; Groman et al., 2019), enhanced learning from negative feedback (Rudebeck and Murray, 2008; Izquierdo et al., 2013; Taswell et al., 2021), and even improvements of deficits produced by OFC lesions (Stalnaker et al., 2007). Yet BLA has not been extensively studied in the context of flexible reversal learning of stimuli vs. actions with the exception of a recent lesion study in rhesus macaques (Taswell et al., 2021). BLA has also not been systematically evaluated for its contributions to deterministic vs. probabilistic schedules, with the exception of another lesion study in monkeys (Costa et al., 2016). The idea that BLA encodes changes in the environment in terms of salience and

98 associability (Roesch et al., 2010) suggests this region may facilitate rapid updating to incorporate
99 new information. The contribution of BLA to reversal learning and its dependence on the nature
100 of the association (i.e., stimulus- vs. action-based), sensory modality (i.e., visual), and type of
101 uncertainty introduced by the task design (i.e., deterministic vs. probabilistic, but also first reversal
102 versus all subsequent reversals) has not been extensively studied using a chemogenetic approach
103 in rats.

104 In parallel, studies with manipulations in rat OFC in reversal learning have included targeting
105 of the entire ventral surface (Izquierdo, 2017), or systematic comparisons of medial vs. lateral OFC
106 (Hervig et al., 2020; Verharen et al., 2020). Here we examined the role of vlOFC, a subregion not
107 as often probed in reward learning as medial and more (dorso)lateral OFC (cf. Zimmermann et al.
108 (2018)) but also densely-interconnected with BLA (Barreiros et al., 2021a; Barreiros et al., 2021b).
109 Additionally, unlike almost all previous studies on reversal learning, we included both male and
110 female subjects.

111 Using a within-subject counterbalanced design, we inactivated these regions prior to reversal
112 sessions and measured both learning and detection/adjustment to reversals. We found that vlOFC,
113 but not BLA, inhibition impaired detection of deterministic and probabilistic stimulus-based,
114 reversals. Conversely, BLA and vlOFC inhibition resulted in significantly slower action-based
115 reversal learning in females, but not males, suggesting a sex-dependent role for these regions.
116 Learning in females was more impacted in first (deterministic) reversal by vlOFC inhibition, and
117 more robustly affected by BLA inhibition in the subsequent probabilistic reversal. These results
118 suggest similar roles for vlOFC and BLA in flexible action-based learning, but a more specialized
119 role for vlOFC in setting adjustments in stimulus-based learning. Further, fitting choice data with
120 reinforcement learning models indicated the attenuated probabilistic action-based reversal learning

deficits were mediated by a larger memory decay for the unchosen option, especially following vlOFC inhibition. Finally, our findings underscore the importance of including male and female animals in neuroscience studies, adding to mounting evidence of sex-modulated learning (Chen et al., 2021a; Chen et al., 2021b).

Materials and Methods

Subjects

Animals for behavioral experiments were adult (N=56, 25 females; 52 used for behavioral study and 4 for *ex vivo* imaging) Long-Evans rats (Charles River Laboratories) average age post-natal-day (PND) 65 at the start of experiments, with a 280g body weight minimum for males and 240g body weight minimum for females at the time of surgery and the start of the experiment. Rats were approximately PND 100 (emerging adulthood; Ghasemi et al. (2021)) when behavioral testing commenced. Before any treatment, all rats underwent a 3-day acclimation period during which they were pair-housed and given food and water *ad libitum*. During that time, they remained in their home cage with no experimenter interference. Following this 3-day acclimation period, animals were handled for 10 min per animal for 5 consecutive days. During the handling period, the animals were also provided food and water *ad libitum*. After the handling period, animals were individually-housed under standard housing conditions (room temperature 22–24° C) with a standard 12 h light/dark cycle (lights on at 6am). Animals were then surgerized and tested on discrimination and reversal learning 1-week post-surgery. At the point of reversal, they were beyond the 3-week expression time for Designer Receptors Exclusively Activated by Designer Drugs (DREADDs).

A separate group of Long-Evans rats (N=4, all males) were used for validation of effectiveness of DREADDs in slides of BLA and vlOFC, using *ex vivo* calcium imaging procedures. All procedures were conducted in accordance to the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and with the approval of the Chancellor's Animal Research Committee at the University of California, Los Angeles.

Surgery

Viral Constructs

Rats were singly-housed and remained in home cages for 4 weeks prior to testing while the inhibitory hM4Di DREADDs expressed in BLA (n=16, 9 females), vlOFC (n=19, 9 females), or eGFP control virus (n=17, 7 females) in these regions. In rats tested on behavior, an adeno-associated virus AAV8 driving the hM4Di-mCherry sequence under the CaMKIIa promoter was used to express DREADDs bilaterally in BLA neurons (0.1 µl, AP = -2.5; ML= ±5; DV = −7.8 and 0.2 µl, AP= -2.5; ML= ±5; DV= −8.1, from bregma at a rate of 0.1 µl/min; AAV8-CaMKIIa-hM4D(Gi)-mCherry, Addgene, viral prep #50477-AAV8). In other animals, this same virus (AAV8-CaMKIIa-hM4Di-mCherry, Addgene) was bilaterally infused into two sites in vlOFC (0.2 µl, AP = +3.7; ML= ±2.5; DV = −4.6 and 0.15 µl, AP= 4; ML= ±2.5; DV= −4.4, from bregma at a rate of 0.1 µl/min). A virus lacking the hM4Di DREADD gene and only containing the green fluorescent tag eGFP (AAV8-CaMKIIa-eGFP, Addgene) was also infused bilaterally into either BLA (n=7), vlOFC (n=5), or anterior cingulate cortex [(n=5); 0.3 µl, AP = +3.7; ML= ±2.5; DV = −4.6, rate of 0.1 µl/min] as null virus controls. Our vlOFC targeting is most similar to infusion sites reported previously by Dalton et al. (2016), and 0.7 mm more medial than others (Costa et

al., 2023). In rats used for *ex vivo* calcium imaging, the same target regions were infused with either GCaMP6f (AAV9-CaMKIIa-GCaMP6f, Addgene), a 1:1 combination of GCaMP6f+mCherry (AAV8-CamKIIa-mCherry, Vector BioLabs, #VB1947), or a 1:1 combination of GCaMP6f+hM4Di-mCherry (same as used for behavior, AAV8-CaMKIIa-hM4Di-mCherry, Addgene).

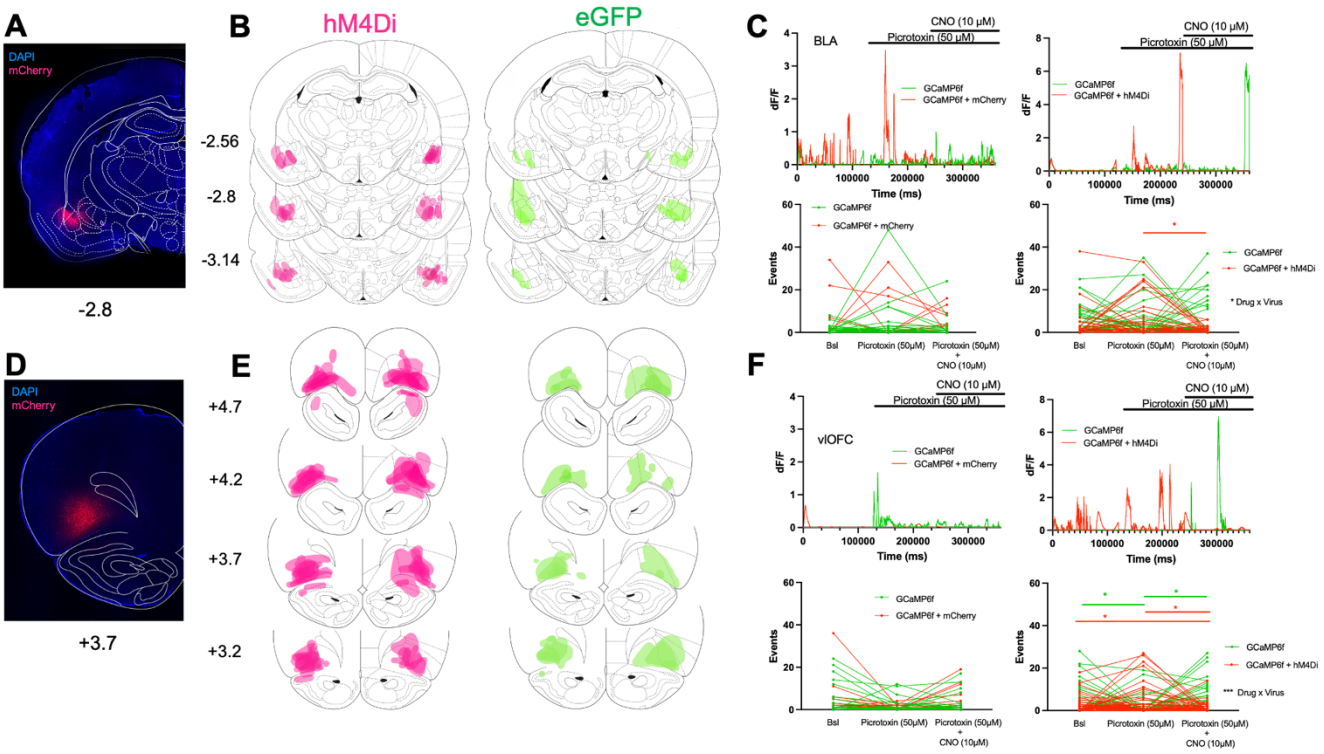

Figure 1. Bilateral targeting of basolateral amygdala (BLA) and ventrolateral orbitofrontal cortex (vlOFC) and confirmation of effective DREADDs inhibition using ex vivo Ca^{2+} imaging in slices. (A) *Photomicrograph of hM4Di-mCherry DREADDs expression in BLA. Numerals indicate mm anterior to Bregma. (**B**) Reconstructions of viral expression of hM4Di (magenta) and enhanced green fluorescent protein, eGFP (green) in BLA. The more intense colors represent regions where expression overlapped the most across animals. (**C**) In BLA neurons expressing GCaMP6f and GCaMP6f+mCherry, application of CNO (10μM) in the presence of picrotoxin (50 μM) had no effect on the frequency of elicited Ca^{2+} events (top left and right: example of single cell traces, bottom left: Ca^{2+} event changes, each line is a cell). In BLA neurons that expressed hM4Di, there was a reduction in the frequency of elicited Ca^{2+} events during CNO application (bottom right). (**D**) Photomicrograph of hM4Di-mCherry DREADDs expression in vlOFC. Numerals indicate mm anterior to Bregma. (**E**) Reconstruction of viral expression of hM4Di (magenta) and enhanced green fluorescent protein, eGFP (green). The more intense colors represent regions where expression overlapped the most across animals. (**F**) In vlOFC neurons expressing GCaMP6f and GCaMP6f+mCherry, application of CNO (10μM) had no effect on the frequency of elicited Ca^{2+} events (top left and right: example of single cell traces, bottom left: Ca^{2+} event changes, each line is a cell). In vlOFC neurons that expressed hM4Di, there was a reduction in the frequency of Ca^{2+} events during CNO application (bottom right). n=2-5 slices/rat, 2-way ANOVA and multiple comparison tests ***p<0.001, *p<0.05.*

Surgical Procedure

Infusions of DREADD or eGFP control virus were performed using aseptic stereotaxic techniques under isoflurane gas (1-5% in O_2) anesthesia prior to any behavioral testing experience. Before surgeries were completed, all animals were administered 5 mg/kg s.c. carprofen (NADA #141–199, Pfizer, Inc., Drug Labeler Code: 000069) and 1cc saline. After being placed in the stereotaxic apparatus (David Kopf; model 306041), the scalp was incised and retracted. The skull was leveled with a +/- 0.3 mm tolerance on the A-P to ensure that bregma and lambda were in the same horizontal plane. Small burr holes were drilled in the skull above the infusion target. Virus was bilaterally infused at a rate of 0.01 µl per minute in target regions (coordinates above). After each infusion, 5 min elapsed before exiting the brain.

Histology

At the end of the experiment, rats were euthanized with an overdose of Euthasol (Euthasol, 0.8 mL, 390 mg/mL pentobarbital, 50 mg/mL phenytoin; Virbac, Fort Worth, TX), were transcardially perfused, and their brains removed for histological processing. Brains were fixed in 10% buffered formalin acetate for 24 h followed by 30% sucrose for 5 days. To visualize hM4Di-mCherry and eGFP expression in BLA or vlOFC cell bodies, free-floating 40-µm coronal sections were mounted onto slides and cover-slipped with mounting medium for DAPI. Slices were visualized using a BZ-X710 microscope (Keyence, Itasca, IL), and analyzed with BZ-X Viewer and analysis software.

Reconstructions of viral expressions of hM4Di (magenta) and green fluorescent protein, eGFP (green) across the AP plane (**Fig. 1BE**) were conducted using Photoshop and Illustrator (Adobe, Inc., San Jose, CA) by individuals blind to condition. Two independent raters then used ImageJ (U. S. National Institutes of Health, Bethesda, Maryland, USA) to trace and quantify pixels

196 at AP +3.7 (vlOFC) and AP -2.8 (BLA) for each animal. Three measures were obtained per
197 hemisphere and were highly correlated (Spearman rank correlation: $r = 0.93$, $p < 0.01$). Only
198 subjects with bilateral expression were included in behavioral analyses (3 BLA rats excluded due
199 to unilateral expression). There were no differences in expression level between males and females
200 for pixel count reconstructions [$F_{(1,12)} = 0.32$, $p = 0.58$], but there was a significant difference
201 between target brain region (vlOFC, BLA) in the hM4Di virus group [$F_{(1,12)} = 9.71$, $p = 0.009$]; the
202 latter was expected given the larger infusion volumes in vlOFC.
203
204 **Food Restriction**
205 Five days prior to any behavioral testing, rats were placed on food restriction with females
206 on average maintained on 10-12 grams/ day and males given 12-14 grams/ day of chow. Food
207 restriction level remained unchanged throughout behavioral testing, provided animals completed
208 testing sessions. Water remained freely available in the home cage. Animals were weighed every
209 other day and monitored closely to not fall below 85% of their maximum, free-feeding weight.
210
211 **Drug administration**
212 Inhibition of vlOFC or BLA was achieved by systemic administration of clozapine-N-
213 oxide, CNO (i.p., 3mg/kg in 95% saline, 5% DMSO) in animals with DREADDs. Rats with eGFP
214 in these regions underwent identical drug treatment. Rats were randomly assigned to drug
215 treatment groups, irrespective of performance in pretraining. CNO was administered only during
216 reversal learning, 30 min prior to behavioral testing. We followed previous work on timing and
217 dose of systemic CNO (Stolyarova, Rakhshan et al., 2019; Hart et al., 2020) and considering the
218 long duration of test sessions. To control for nonspecific effects of injections and handling stress,

we also injected animals with saline vehicle (VEH). To increase power and decrease the number of animals used in experiments, we used a within-subject design for assessing the effects of CNO, with all rats receiving CNO and VEH injections in a counterbalanced order. Thus, for drug administration if a rat received CNO on the first reversal (R1), it was administered VEH on the second reversal (R2), CNO on the third reversal (R3), and VEH on the fourth reversal (R4), or vice versa: VEH on R1, CNO on R2, VEH on R3, and CNO on R4.

Behavioral Testing

Pretraining. Behavioral testing was conducted in operant conditioning chambers outfitted with an LCD touchscreen opposing the sugar pellet dispenser. All chamber equipment was controlled by customized ABET II TOUCH software (Lafayette Instrument Co., Lafayette, IN).

The pretraining protocol, adapted from established procedures (Stolyarova and Izquierdo, 2017), consisted of a series of phases: Habituation, Initiation Touch to Center Training (ITCT), Immediate Reward Training (IMT), designed to train rats to nose poke, initiate a trial, and select a stimulus to obtain a reward (i.e., sucrose pellet). Pretraining stages have been reported in detail elsewhere (Stolyarova, Rakhshan et al., 2019). For habituation pretraining, the criterion for advancement was collection of all 5 sucrose pellets. For ITCT, the criterion to the next stage was set to 60 rewards consumed in 45 min. The criterion for IMT was set to 60 rewards consumed in 45 min across two consecutive days. After completion of all pretraining schedules, rats were advanced to the discrimination (initial) phase of either the action- or stimulus-based reversal learning task, with the task order counterbalanced (**Fig. 2AB** or **Fig. 4AB**). A subset of animals was tested first on the action-based task (10 vlOFC hM4Di, 9 BLA hM4Di), while others were

tested first on the stimulus-based task (9 vlOFC hM4Di, 7 BLA hM4Di, 17 eGFP). Three vlOFC rats completed only the stimulus-based task.

Action-based deterministic discrimination learning. After completion of either all pretraining schedules or all four reversals of the stimulus-based task, rats were advanced to the discrimination (initial) phase of the action-based task (**Fig. 2A**). Rats were required to initiate a trial by touching the white graphic stimulus in the center screen (displayed for 40 s), and after initiation rats would be presented with two stimuli (i.e., fan or marble) on the left and right side of the screen (displayed for 60 s). Rats could nosepoke either the spatial side rewarded with one sucrose pellet (the better option, $p_R(B)=1.0$;) or the spatial side that went unrewarded (the worse option, $p_R(W)=0.0$). Thus, rats were required to ignore the properties of the stimuli and determine the better rewarded side. If a side was not selected, it was scored as a choice omission, and a 10 s inter-trial interval (ITI) ensued. If a trial was not rewarded, a 5 s time-out would occur, followed by a 10 s ITI. If a trial was rewarded, a 10 s ITI would occur after the reward was collected. The criterion was set to 60 or more rewards consumed and selection of the correct option in 75% of the trials or higher during a 60 min session across two consecutive days. After reaching the criterion for the discrimination phase, the rats were advanced to the reversal phase beginning on the next session. Animals were not administered either CNO or VEH injections during discrimination learning.

Action-based reversal learning. After the discrimination phase, the rats advanced to the reversal phase. Before a reversal learning session, rats were injected intraperitoneally with either 3 mg/kg of CNO or VEH 30 min prior to each reversal testing session. The side previously associated with the $p_R(B)=1.0$ probability was now associated with a $p_R(W)=0.0$ probability of being rewarded, and vice versa. The criterion was the same as the deterministic discrimination

phase. After reaching the criterion for the first deterministic reversal phase (i.e., R1), the rats advanced to the second deterministic reversal phase (i.e., R2) beginning on the next session. Rats that had previously received VEH during the first reversal would now receive CNO injections, and vice versa.

After completing both deterministic reversal learning phases, rats advanced to the first probabilistic reversal learning phase (i.e., reversal 3, R3). Rats underwent the same injection procedure as the prior reversals. However, the spatial side (i.e., left or right) previously associated with $p_R(B)=1.0$, was now associated with a $p_R(W)=0.1$ probability of being rewarded, whereas the spatial side previously associated with $p_R(W)=0.0$ probability, was now associated with $p_R(B)=0.9$. The criterion was the same as the previous deterministic reversal learning phases. After reaching the criterion for the first probabilistic reversal learning phase (i.e., reversal 3, R3), rats were advanced to the second probabilistic reversal phase (i.e., reversal 4, R4) beginning on the next testing day, where the probabilities would be reversed once again. Rats that had previously received VEH during the first probabilistic reversal now received CNO injections, and vice versa.

Stimulus-based deterministic discrimination learning. After completion of all pretraining schedules (or all reversals of the action-based task), rats were advanced to the discrimination (initial) phase of learning in which they would initiate a trial by touching a white graphic stimulus in the center screen (displayed for 40 s), and choose between two different visual stimuli pseudorandomly presented on the left and right side of the screen. Stimuli were displayed for 60 s each, randomly assigned as the better or worse stimulus: $p_R(B)=1.0$ or $p_R(W)=0.0$. If a trial was not initiated within 40 s, it was scored as an initiation omission. If a stimulus was not selected, it was scored as a choice omission, and a 10 s ITI ensued. If a trial was not rewarded, a 5 s time-out would occur, followed by a 10 s ITI. Finally, if a trial was rewarded, a 10 s ITI would follow after

the reward was collected. A criterion was set to 60 or more rewards consumed and selection of the correct option in 75% of the trials or higher during a 60 min session across two consecutive days. After reaching the criterion for the discrimination phase or if rats were unable to achieve criterion after 10 days, rats were advanced to the reversal phase beginning on the next session. Animals were not administered either CNO or VEH injections during discrimination learning.

Stimulus-based reversal learning. After the discrimination phase, rats advanced to the first deterministic reversal learning phase (i.e., reversal 1, R1) where they were required to remap stimulus-reward contingencies. As above, before a reversal learning session, rats were injected intraperitoneally with either 3 mg/kg of CNO or VEH 30 min prior to each reversal testing session. The criterion was the same as discrimination learning. After reaching the criterion for the first reversal phase or if they were unable to achieve criterion after 10 days, the rats were advanced to the second deterministic reversal phase (i.e., reversal 2, R2) beginning on the next testing day, where the reward contingencies were reversed once again. Rats that had previously received VEH during the first reversal now received CNO injections, and vice versa.

After completing both deterministic reversal learning phases, rats advanced to the first probabilistic reversal learning phase (i.e., reversal 3, R3). The injection procedure remained the same as prior reversals. However, the visual stimulus previously associated with $p_R(B)=1.0$, would now be associated with $p_R(W)=0.1$, whereas the stimulus previously associated with $p_R(W)=0.0$, would now be associated with $p_R(B)=0.9$. The criterion remained the same as prior reversals. After reaching the criterion for the first probabilistic reversal learning phase or if rats were unable to achieve criterion after 10 days, the rats were advanced to the second probabilistic reversal phase (i.e., reversal 4, R4) beginning on the next testing day, where the probabilities would be reversed

once again. As above for action-based reversal learning, rats that had previously received VEH during the first probabilistic reversal now received CNO injections, and vice versa.

Ex vivo calcium imaging

In N=4 animals (all males), following >3 weeks following stereotaxic viral injections, rats (n=1 rat/brain region/virus combination; n=2-5 slices/rat) were deeply anesthetized with isoflurane (Patterson Veterinary, MA, USA), decapitated and brains submerged in ice cold oxygenated (95/5% O_2/CO_2) slicing artificial cerebrospinal fluid (ACSF) containing (in mM): 62 NaCl, 3.5 KCl, 1.25 NaH_2PO_4, 62 choline chloride, 0.5 $CaCl_2$, 3.5 $MgCl_2$, 26 $NaHCO_3$, 5 N-acetyl L-cysteine, and 5 glucose, pH adjusted to 7.3 with KOH. Acutely microdissected vlOFC or BLA slices (300 μM thick) were obtained (VT1200s, Leica, Buffalo Grove, IL) and transferred to room temperature normal ACSF containing (in mM): 125 NaCl, 2.5 KCl, 1.25 NaH_2PO_4, 2 $CaCl_2$, 2 $MgCl_2$, 26 $NaHCO_3$, 10 glucose, pH adjusted to 7.3 with KOH, and allowed equilibrate for >1 hr prior to transfer into a perfusion chamber for imaging.

Imaging was performed on a Scientifica SliceScope, with imaging components built on an Olympus BX51 upright fluorescence microscope equipped with an sCMOS camera (Hamamatsu Orca Flash 4.0v3). Anatomical regions in brain sections for Ca^{2+} imaging were first identified by brightfield imaging with 780nm LED (Scientifica) illumination. Ca^{2+} imaging was performed using a 40x, 0.80NA water immersion objective (Olympus), continuous 470nm LED illumination (ThorLabs), and a filter cube suitable for GCaMP6f imaging: Excitation: Brightline 466/40, Dichroic: Semrock FF495-Di03, Emission: Brightline 525/50. Slices were housed on poly-D lysine cover slips attached to RC-26G chamber (Warner Instruments, Holliston, MA), which was modified with platinum wires to apply electric field stimulation. Images were acquired continually with 20 ms exposure time. Electric field stimulation was applied at 110 mV (twin pulse every 5s).

Temperature of ACSF during the recorded sessions was held at 28° C to minimize bubble formation.

Calcium data extraction

Prior to imaging sessions, 40x images of red and green fluorescence were captured and subsequently overlaid for post-hoc genotyping of individual cells (GCaMP6f^{+}, GCaMP6f^{+}/hM4Di-mCherry^{+}, or GCaMP6f/mCherry^{+}). Blinded scorers semi-manually curated regions of interest (ROIs) using Python-based Suite2P software (Pachitariu et al., 2017). ROI fluorescence was subtracted from the annular surround fluorescence, low-pass filtered, and transformed to dF/F_0 as previously described (Asrican and Song, 2021) where F_0 is calculated with a boxcar filter with a 200-frame lookback window. dF/F_0 values were clipped between 0 and 9000 to eliminate negative changes. Area under the curve and event frequency of each cell was calculated for each drug treatment. A threshold of 0.15 dF/F was used to determine significant events, which is lower than the dF/F of a single *ex vivo* action potential, but significantly above signal to noise in our recorded traces (Tada et al., 2014), **Fig. 1CF**.

Data Analyses

MATLAB (MathWorks, Natick, Massachusetts; Version R2021a) was used for all statistical analyses and figure preparation. Data were analyzed with a series of mixed-effects General Linear Models (GLM) for the discrimination learning phase to establish there were no baseline differences in learning measures between the hM4Di and eGFP animals (i.e., virus group) prior to any drug treatment for each task separately. Mixed-effects GLMs were also conducted on reversal phases, with all fixed factors included in the model [i.e., reversal number (1-4), virus

group (hM4Di, eGFP), drug (CNO, VEH), sex (female, male), drug order (CNO1, VEH1)] and individual rat as a random factor. These GLMs were run for each task type (stimulus- and action-based tasks) separately. Since learning reached asymptote at 5-days for stimulus-based reversal learning, only the first 5 days were included in the GLM. Similarly, since rats typically reached a plateau (and criterion) at 150 trials for action-based reversal learning, we included only the first 150 trials in the GLM. Significant reversal number and/or drug order interactions were further analyzed with a narrower set of fixed factors and Bonferroni-corrected post-hoc comparisons. In the instance where sex was found a significant predictor (moderator), sex was entered as a covariate factor in subsequent reversals. Accuracy (probability correct) before and after a reversal (-3 and +3 sessions surrounding a reversal) was analyzed using ANOVA with virus group (hM4Di, eGFP) and drug order (CNO1, VEH1) as fixed factors on the average change pre-post reversal. Virus expression level was analyzed with ANOVA by sex (male, female) and region (vlOFC, BLA) on pixel counts.

Dependent measures for learning included probability of choosing the correct or better option, initiation latencies and omissions (failure to initiate a trial, latency to initiate a trial, respectively), correct and incorrect choice latencies (latency to select the correct or better stimulus or spatial side and latency to select the incorrect or wrong stimulus or spatial side, respectively), reward latencies (latency to collect the reward), probability of win-stay, and probability of lose-shift. Probability of win-stay and lose-shift adaptive strategies were calculated for the stimulus-based task such that each trial was classified as a *win* if an animal received a sucrose pellet, and as a *loss* if no reward was delivered. Statistical significance was noted when p-values were less than 0.05. All Bonferroni post-hoc tests were corrected for number of comparisons.

To analyze *ex vivo* calcium imaging data, 2-way ANOVAs with drug and virus as factors were conducted to compare calcium event changes in GCaMP6f and GCaMP6f+mCherry in each brain region for control experiments, and for GCaMP6f and GCaMP6f+hM4Di-mCherry in each brain region for the experimental group. Tests corrected for number of comparisons were conducted for interactions.

Reinforcement Learning Models

To capture the differences in learning and choice behavior during the action-based task, we utilized two conventional reinforcement learning (RL) models. Specifically, the subjective estimate of reward (V) for each choice option was updated on a trial-by-trial basis using reward prediction error (RPE), the discrepancy between actual and expected reward value. In the first model, which we refer to as *RL*, the value estimate of the chosen option (V_C) for a trial t was updated using the following equations:

$$V_C(t+1) = V_C(t) + \alpha\big(R(t) - V_C(t)\big), \qquad (1)$$

where $R(t)$ indicates the presence (1) or absence (0) of a reward for the given trial, and α is the learning rate dictating the amount of update in the value estimate by RPE. In this model, the value of unchosen option was not updated.

The second model, referred to as RL_{decay}, used the same learning rule as Equation (1) for updating the value of the chosen option, and additionally updated the value of the unchosen option (V_U) as follows:

$$V_U(t+1) = V_U(t) - \gamma_d(V_U(t)), \qquad (2)$$

where γ_d is a decay rate controlling the amount of passive decay in value of the unchosen option. In both models above, the probability of choosing a particular option was computed using the following decision rule:

$$P_i(t) = (1 + e^{-\beta\left(V_i(t) - V_j(t)\right)})^{-1}, \quad (3)$$

where *i* and *j* corresponds to two alternative options (i.e., left and right for action-based task), and β is the inverse temperature or sensitivity governing the extent to which higher-valued options are consistently selected.

We used the standard maximum likelihood estimation method to fit choice data and estimate the parameters for each session of the experiment. The values of the learning rate α and decay rate γ_d were bounded between 0 and 1, and β was bounded between 1 and 100. Initial parameter values were selected from this range, and fitting was performed using the MATLAB function *fmincon*. For each set of parameters fitted to each session, we repeated 30 different initial conditions selected from evenly-spaced search space to avoid local minima. The best fit was selected from the iteration with the minimum negative log-likelihood (*LL*). For the first model (*RL*), we treated the uninitiated or uncommitted trials with no choice data as if they had not occurred. In contrast, for the second model (RL_{decay}), both choice options were considered unchosen for those trials and both of the value estimates decayed passively according to Equation (2).

To quantify goodness of fit, we computed both Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for each session as follows:

$$AIC = -2 * LL + 2 * k, \quad (4)$$

$$BIC = -2 * LL + \ln(n) * k, \quad (5)$$

where k is the number of parameters in the model (two for RL and three for RL_{decay}), and n is the number of choice trials in the session.

Results

Ex vivo calcium imaging in slices

We performed ex *vivo* Ca^{2+} imaging to confirm the selective action on $CaMKII^{+}$ neuronal excitability in vlOFC and BLA in rats expressing hM4Di DREADD vs. controls expressing mCherry. In BLA, there was no significant effect of CNO (10μM) on Ca^{2+} events for neurons expressing GCaMP6f or GCaMP6f+mCherry (**Fig. 1C**). A 2-way ANOVA resulted in a significant drug × virus interaction [$F_{(2,324)} = 3.367$, $p = 0.036$], with a selective reduction in the frequency of elicited Ca^{2+} events during CNO only in neurons expressing GCaMP6f+hM4Di (multiple comparison test, p=0.049).

In vlOFC, there was also no significant effect of CNO (10μM) on Ca^{2+} events for neurons expressing GCaMP6f or GCaMP6f+mCherry (**Fig. 1F**). However, in $CaMKII^{+}$ vlOFC neurons expressing GCaMP6f+hM4Di there was a decrease in the frequency of Ca^{2+} events during CNO application. A 2-way ANOVA revealed a significant drug x virus interaction [$F_{(2,400)} = 8.349$, $p < 0.001$], with multiple comparisons test resulting in decreased Ca^{2+} events in GCaMP6f+hM4Di following CNO ($p = 0.02$), and increased activity in GCaMP6f expressing neurons after CNO ($p = 0.02$).

Discrimination learning: eGFP controls

Mixed-effects GLMs for the discrimination learning phase were conducted for each task separately to establish if there were baseline differences in learning measures between animals

444 infused with eGFP virus in different brain regions. There were no differences between the eGFP
445 groups by target region on learning (i.e., the probability of choosing the correct side) across trials
446 in the action-based task (β_{region} = -0.13, t(2392) = -0.72, p = 0.47), as well as no differences in
447 learning (i.e., probability of choosing the correct visual stimulus) across sessions in the stimulus-
448 based task (β_{region} = 0.10, t(77) = 1.70, p = 0.09). Thus, animals' data were collapsed into a single
449 eGFP virus group for subsequent analyses.
450
451 ***Discrimination learning: hM4Di vs. eGFP***
452 For the action-based task, there were no significant effects of virus or virus interactions for
453 vlOFC vs. eGFP on *probability correct* (β_{virus} = -0.13, t(4792) = -1.02, p = 0.31), with similar
454 findings for the comparison of BLA vs. eGFP (β_{virus} = -0.14, t(4792) =- 1.06, p= 0.29; **Fig. 2C**).
455 All animals met criterion very quickly (~2 days), thus, we compared trials to reach 75% criterion
456 (i.e., probability of choosing the correct side). Both hM4Di virus groups performed comparably
457 [M±SEM: vlOFC hM4Di (81.1±23.0), BLA hM4Di (84.1±21.7)], whereas the eGFP group met
458 criterion within fewer trials (59.5±18.6), but the difference was not statistically significant [vlOFC
459 hM4Di vs. eGFP: β_{virus} = 54.5, t(25) = 1.34, p = 0.19; BLA hM4Di vs. eGFP: β_{virus} = 54.8, t(25) =
460 1.38, p = 0.18].
461 For the stimulus-based task, there were also no significant effects of virus or virus
462 interactions for either vlOFC vs. eGFP on *probability correct* (β_{virus} = -0.06, t(141) = -1.26, p =
463 0.21), or for BLA vs. eGFP (β_{virus} = -0.05, t(152) = -1.32, p = 0.19; **Fig. 4C**). The animals on
464 average took approximately ~6 days to meet criterion regardless of virus group [M±SEM: vlOFC
465 hM4Di (6.1±0.7), BLA hM4Di (6.5±1.2), eGFP (6.9±0.7)]. However, many animals did not meet
466 criterion after a maximum of 10 days of testing (61% of rats).

Given the poorer learning in the stimulus-based task, we evaluated whether this was due to the order of task administered [i.e., Stimulus → Action or Action → Stimulus]. To test whether learning was influenced by task order, we analyzed *probability correct* during initial discrimination learning for the stimulus-based task, which resulted in no effect of task order (β_{order} = 0.03, t(220) = 0.97, p = 0.33), but a significant task order x session interaction ($\beta_{order\ x\ session}$ =-0.03, t(220) = -3.06, p = 0.002). Thus, subsequent analyses were conducted with task order analyzed separately by session, which revealed that animals administered the Action → Stimulus task order exhibited poorer learning across sessions ($\beta_{session}$ = 0.01, t(58) = 2.03, p = 0.05), compared to those administered the Stimulus→ Action task order ($\beta_{session}$ = 0.04, t(162) = 6.14, p < 0.0001). Notably, only 1 BLA hM4Di animal successfully met criterion on stimulus-based learning for the Action → Stimulus order, while no OFC hM4Di animals achieved this.

Accuracy in reversal learning across session and trials

Action-based reversal learning. Mixed-effects GLMs were used to analyze *probability correct*, with trial number, reversal number, drug order, drug, virus, and sex as between-subject factors, trial number, reversal number, and drug as within-subject factors, and individual rat as random factor. GLMs were conducted separately by target region (vlOFC vs. eGFP and BLA vs. eGFP), using the following formula for the full model: $\gamma \sim [1 + trial\ number * reversal\ number * virus * drug * drug\ order * sex + (1 + trial\ number * reversal\ number * drug | rat)]$.

For the comparison of vlOFC with eGFP, several interactions were found: interaction of sex, virus, drug, and trial number ($\beta_{sex\ x\ virus\ x\ drug\ x\ trial\ number}$ = 0.035, t(19136) = 2.16, p = 0.03), an interaction of sex, virus, drug, drug order, and trial number ($\beta_{sex\ x\ virus\ x\ drug\ order\ x\ trial\ number}$ = -0.006, t(19136) = -2.32, p = 0.02), as well as an interaction of virus, drug, drug order, reversal number,

and trial number ($\beta_{\text{virus x drug x drug order x reversal number x trial number}}$ = -0.001, t(19136) = -2.19, p = 0.03). Due to these interactions, we were justified to look further at individual reversals. For the comparison of vlOFC hM4Di vs eGFP, sex emerged as a significant predictor of R1 probability correct (β_{sex} = -0.279, t(4784) = -2.03, p = 0.04), and thus sex was entered in the model as a covariate: *γ ~ [1 + trial number * reversal number * virus * drug order + sex + (1 + trial number * drug| rat)*]. With sex as a covariate, there was a significant effect of trial number ($\beta_{\text{trial number}}$ = 0.003, t(4791) = 7.08, p = 1.63e^{-12}) and a nonsignificant but trending interaction of virus, drug order, and trial number for R1: ($\beta_{\text{virus x drug order x trial number}}$ = -0.002, t(4791) = -1.80, p = 0.07; **Fig. 2D**). For probability correct in R3, we also found sex a significant moderator, interacting with drug ($\beta_{\text{sex x drug}}$ = 0.334, t(4784) = 2.18, p = 0.029), and drug x trial number ($\beta_{\text{sex x drug x trial number}}$= -0.003, t(4784) = -2.55, p = 0.01; **Fig. 2D**). When sex was entered as a covariate in the model, we found a

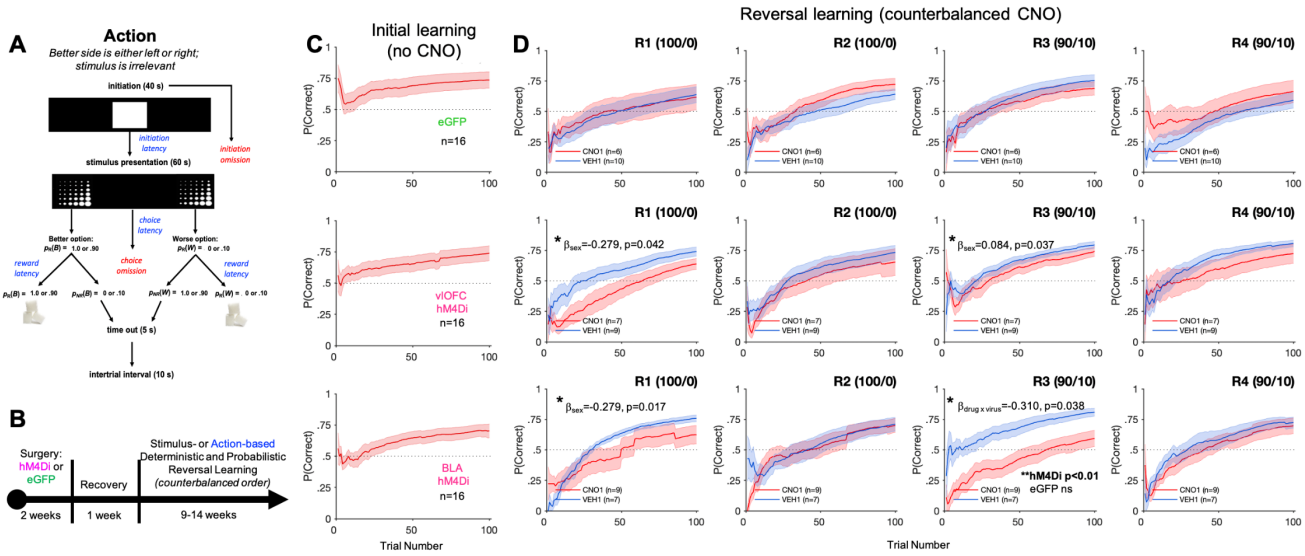

Figure 2. Either BLA or vlOFC inhibition attenuates action-based reversal learning. (A-B) *Trial structure (A) and timeline (B) of the action-based task. Rats were first surgerized with either hM4Di DREADDs on a CaMKII promoter or eGFP null virus, also on the same promoter. Rats were allowed to recover for 1 week before testing on a stimulus-or action-based reversal learning task.* ***(C)*** *Initial learning of a rewarded side.* ***(D)*** *Learning during subsequent deterministic (100/0) and probabilistic (90/10) reversals. Plots show cumulative P(Correct) for first 100 trials with a sliding window of 10 trials is shown. Drug order was counterbalanced such that on R2 and R4 animals received VEH if they were administered CNO first on R1 and R3, and vice versa. There was no effect of CNO on learning in the eGFP group. *p<0.05 sex as a significant predictor (see* ***Fig. 3*** *for learning curves plotted by sex). Bonferroni-corrected post-hocs following mixed-effects GLM with sex as a covariate fixed factor wherein a drug x virus interaction was found resulted in **p<0.01 effect of drug only in BLA hM4Di, not in eGFP.*

significant effect of trial number as all animals exhibited improvements in probability correct across trials ($\beta_{\text{trial number}} = 0.003$, $t(4791) = 7.73$, $p = 1.34e^{-14}$), with males performing significantly better than females (β_{sex} 0.0841, $t(4791) = 2.088$, $p = 0.04$; **Fig. 3B**), but no other significant interactions with any other factor for action-based reversal learning for vlOFC hM4Di compared to eGFP.

Starting with the full model as above for the comparison of BLA with eGFP, several interactions of virus x drug were observed, including: virus x drug order x trial number, ($\beta_{\text{virus x drug order x trial number}} = -0.008$, $t(18986) = -2.58$, $p = 0.01$), virus x reversal number x drug order x trial number ($\beta_{\text{virus x reversal number x drug order x trial number}} = 0.002$, $t(18986) = 2.59$, $p = 0.01$), and virus x drug x reversal number x drug order x trial number ($\beta_{\text{virus x drug x reversal number x drug order x trial number}} = -0.002$,

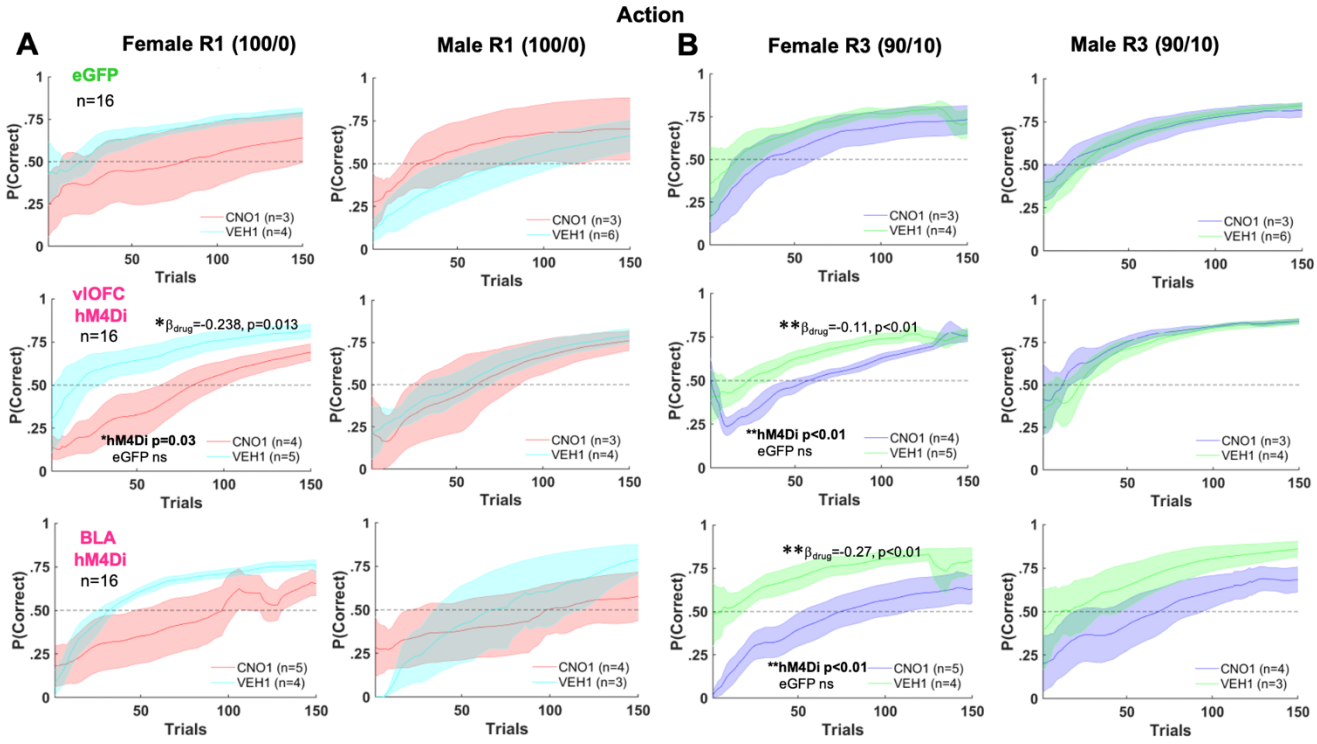

Figure 3. Female learning was more adversely affected by vlOFC and BLA inhibition than male learning during deterministic and probabilistic action-based reversals. *Cumulative P(Correct) for first 150 trials with a sliding window of 10 trials is shown. Learning of deterministic (100/0) (A) and probabilistic (90/10) (B) reversals as measured by probability correct (P(Correct)). Drug order was counterbalanced such that on R2 and R4 (not shown) animals received VEH if they were administered CNO first on R1 and R3, and vice versa. Chemogenetic inhibition of vlOFC lowered P(Correct) on first deterministic R1 and first probabilistic reversal R3, whereas BLA inhibition attenuated probabilistic learning. Bonferroni-corrected post-hocs following GLM following a sex x drug interaction resulted in effect of drug only in vlOFC or BLA hM4Di, not in eGFP. There were no significant sex differences and no effect of CNO on learning in the eGFP group. **p<0.01, *p<0.05.*

t(18986) = -2.00, p = 0.046). There were also interactions with sex, including: sex x drug x drug order ($\beta_{\text{sex x drug x drug order}}$ = -1.07, t(18986) = -2.72, p = 0.01), and sex x drug x reversal number x drug order ($\beta_{\text{sex x drug x reversal number x drug order}}$ = 0.335, t(18986) = 2.19, p = 0.03). Due to these interactions, we were justified to look further at individual reversals. For BLA hM4Di compared to eGFP in R1 there was also a sex difference (females > males) and an effect of trial number on probability correct (β_{sex} = -0.279, t(4784) = -2.38, p = 0.02; $\beta_{\text{trial number}}$ = 0.002, t(4784) = 3.87, p = $1.0e^{-04}$). With sex as a covariate, there was only a significant effect of trial number for R1 ($\beta_{\text{trial number}}$ = 0.003, t(4791) = 7.17, p = $8.93e^{-13}$). Sex was also a significant moderator of probability correct in R3, with significant interactions of sex x trial number ($\beta_{\text{sex x trial number}}$= 0.002, t(4784) = 2.25, p = 0.03) and sex x trial number x drug ($\beta_{\text{sex x drug x trial number}}$= -0.003, t(4784) = -2.42, p = 0.02). When sex was included as a covariate, there was a significant effect of BLA inhibition on probability correct in probabilistic R3: (GLM: $\beta_{\text{drug*virus}}$ = -0.31, t(4791)= -2.08, p = 0.038). Bonferroni-corrected post-hoc comparisons revealed an effect of CNO in hM4Di (p < 0.01), not in eGFP (p = 1.0), with both males and females exhibit attenuated learning of probabilistic R3 following BLA inhibition.

Given the sex x drug interactions observed in both hM4Di groups, *probability correct* was also analyzed separately for hM4Di males, eGFP males, hM4Di females, and eGFP females using the following formula: $\gamma \sim [1 + drug + (1 + drug| rat)]$. We found a significant effect of drug for R1 (β_{drug}= -0.24, t(1348) = -2.49, Bonferroni corrected p = 0.03) and R3 (β_{drug}= -0.11, t(1348) = -3.09, Bonferroni corrected p = 0.004) for vlOFC hM4Di females and significant effect of drug for R3 for BLA hM4Di females (β_{drug}= -0.27, t(1348) = -2.97, Bonferroni corrected p = 0.006). There was no significant effect of drug in eGFP females, eGFP males, or hM4Di males across regions and reversals (**Fig. 3**).

534 To gain more insight into this sex difference and its potential underlying mechanism, we
535 next compared the estimated model parameters from the reinforcement learning models (*RL*,
536 RL_{decay}). Comparing the goodness of fit between the two models, we found that the second model
537 with the decay parameter (RL_{decay}) better accounted for the animals' choice behavior as indicated
538 by significantly lower AIC (paired sample t-test; t(1554) = 6.792, $p = 1.56e^{-11}$). Overall mean BIC
539 value was also lower for the second model, although the values did not significantly differ from
540 the first model (t(1554) = 0.93806, $p = 0.348$). Therefore, we focused on the estimated parameters
541 from the $RL1_{decay}$ model only.

542 Comparison of the estimated parameters across groups revealed that female and male rats
543 differed mainly in the decay parameter γ_d, which governs the amount of passive decay or
544 forgetting in the value estimate of the unchosen option (**Fig. 3-1**). During the first deterministic
545 reversal (R1, 100/0), eGFP females showed overall significantly lower values of γ_d than eGFP
546 males (mean difference in γ_d = -0.098; Wilcoxon rank sum test, p=0.00537), suggesting a different
547 mechanism of adjustment to the reversal between females and males. While there was no clear
548 evidence for such sex difference in γ_d for the vlOFC or BLA hM4Di groups during the first
549 deterministic reversal (R1), the first probabilistic reversal (R3, 90/10) instead revealed a sex-
550 specific effect between CNO and VEH groups: vlOFC hM4Di females who were administered
551 CNO to inhibit vlOFC had significantly higher values of γ_d compared to those who received VEH
552 (mean difference in γ_d = 0.150; $p = 0.00757$). In contrast, vlOFC hM4Di males did not show a
553 significant difference in γ_d between CNO and VEH groups (mean difference in γ_d = -0.0753; $p =$
554 0.228). BLA hM4Di groups showed a similar trend in γ_d, with the females exhibiting a larger
555 difference between CNO and VEH groups (mean difference in γ_d = 0.137; $p = 0.0673$) than males
556 (mean difference in γ_d = 0.052; $p = 0.961$). These results based on RL model fitting suggest that

the attenuated probabilistic learning for the female CNO groups is mediated by larger γ_d or decreased memory for the unchosen options after vlOFC and BLA inhibition. Importantly, this significant difference emerged due to hM4Di VEH females exhibiting *enhanced* memory in R3. This is because in this condition, rats had previously received CNO in R2 and thus did not encode the reversal very well, making it easier to return to the R1 contingency (which was the same as in R3). Collectively, this shows that vlOFC, and perhaps secondarily BLA, is necessary for the encoding and retrieval of action-based values.

Stimulus-based reversal learning. In contrast to the acquisition curves that demonstrated learning of the initial visual discrimination (**Fig. 4C**), all animals exhibited difficulty with stimulus-based reversal learning, rarely achieving above 60% after 10 sessions (**Fig. 4-1**), similar to recent reports (Harris, Aguirre et al., 2021; Ye et al., 2023). Here, due to several non-learners, we adhered to the criterion of rats reaching greater than a 50% running window average for the last 100 trials in discrimination, for inclusion in subsequent reversal learning analyses. The following numbers did not meet this criterion and were excluded from these groups: 0 of 13 vlOFC hM4Di, 6 of 15 BLA hM4Di, and 6 of 17 eGFP. As above, GLMs were conducted separately for accuracy (*probability correct*) by target region comparison, but with session instead of trial number as a within-subject factor. Thus, the GLM formula was as follows: $\gamma \sim [1 + session * reversal\ number * virus * drug * drug\ order * sex + (1 + session * reversal\ number * drug | rat)]$.

For vlOFC hM4Di comparison with eGFP, we observed several interactions including: session, drug, and virus ($\beta_{session\ x\ drug\ x\ virus} = -0.096$, $t(536) = -2.83$, $p = 4.81e^{-03}$), session, drug, drug order, and virus ($\beta_{session\ x\ drug\ x\ drug\ order\ x\ virus} = 0.146$, $t(536) = 2.62$, $p = 9.17e^{-03}$) and drug, drug order, and reversal number ($\beta_{drug\ x\ drug\ order\ x\ reversal\ number} = 0.107$, $t(536) = 2.18$, $p = 0.03$). We also observed a sex x reversal number interaction ($\beta_{sex\ x\ reversal\ number} = -0.084$, $t(536) = -2.50$, $p = 0.01$).

Aside from a significant effect of session in R1 ($\beta_{session}$ = 0.03, t(134) = 2.67, p = 8.55e^{-03}) there were no significant predictors of learning with follow-up GLM analyses with sex in the model. With sex as a covariate in the model there was a significant effect of session for R1 and R3 (R1: $\beta_{session}$ = 0.026, t(141) = 3.80, p = 2.16e^{-04}; R3: $\beta_{session}$ = 0.018, t(141) = 2.92, p = 4.11e^{-03}), and a session x virus interaction in R3 ($\beta_{session\ x\ virus}$ = -0.020, t(141) = -2.07, p = 0.04). Bonferroni-corrected post-hoc comparisons revealed an effect of session in eGFP ($p < 0.01$), but not in hM4Di (p = 0.45), indicating that only the eGFP group improved across session in R3 (**Fig. 4-1**).

For BLA hM4Di compared to eGFP, several interactions were observed including interactions of session, drug, and virus ($\beta_{session\ x\ drug\ x\ virus}$ = -0.087, t(575) = -2.25, p = 0.02), drug, drug order, and reversal number ($\beta_{drug\ x\ drug\ order\ x\ reversal\ number}$ = 0.107, t(575) = 2.29, p = 0.02), and sex and reversal number ($\beta_{sex\ x\ reversal\ number}$ = -0.084, t(575) = -2.15, p = 0.03). However, none of the post-hoc GLM analyses resulted in significant predictors of learning with the exception of session for R1 and R3 (R1: $\beta_{session}$ = 0.026, t(151) = 2.35, p = 0.02; R3: $\beta_{session}$ = 0.018, t(151) = 2.67, p = 8.43e^{-03}) when sex was included as a covariate in the model.

Due to the slow stimulus-based reversal learning, we next assessed *probability correct* around reversals (three sessions before and after reversals for stimulus-based learning, and one session before and after reversals for action-based learning) to test for detection and adjustments to reversals.

Accuracy around reversals: Reversal detection

Stimulus-based reversal learning. We analyzed accuracy (*probability correct*) around reversals, as overall stimulus-based reversal learning was modest. ANOVAs with virus and drug order as between subject factors were conducted on the mean change in accuracy between one

reversal and the next. vlOFC hM4Di was significantly different from eGFP for R1-to-R2 [$F_{(1,24)}$= 9.49, $p < 0.01$] and R2-to-R3 [$F_{(1,24)}$= 10.1, $p < 0.01$], but not R3-to-R4 [$F_{(1,24)}$= 2.61, $p = 0.12$], **Fig. 4D**. In contrast, BLA hM4Di was not significantly different from eGFP on changes in accuracy around any of the reversals.

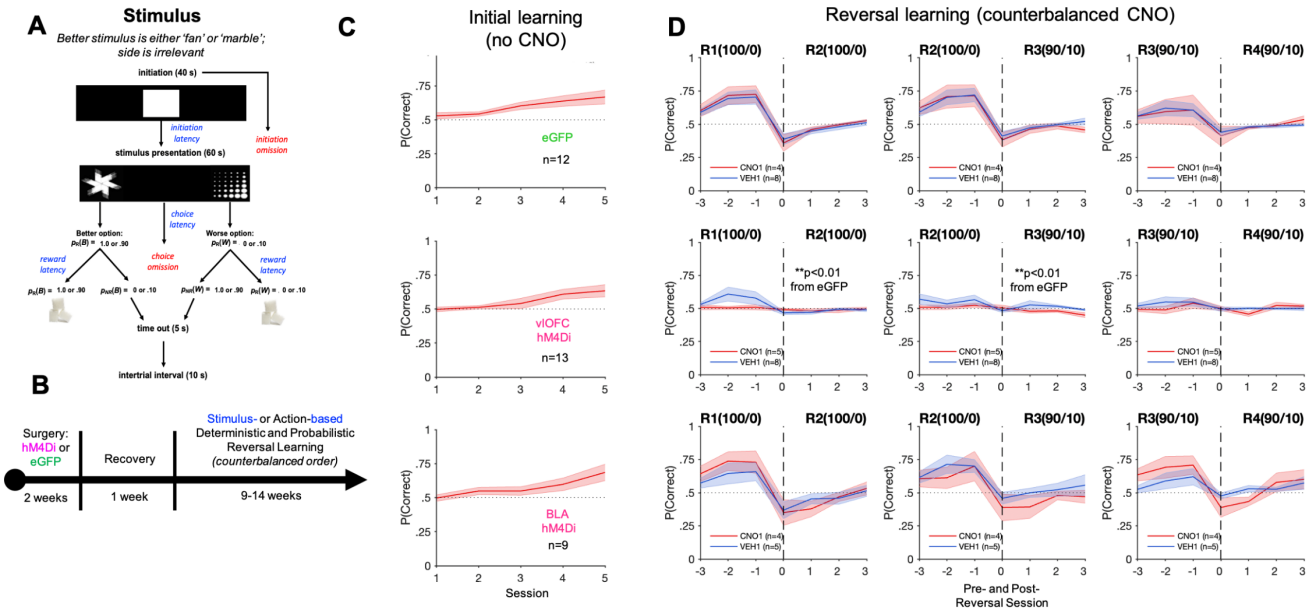

Figure 4. Inhibition of vlOFC, but not BLA, impairs the detection of stimulus-based reversals as measured by probability correct adjustments. (A-B) *Trial structure (A) and timeline (B) of the stimulus-based task. Rats were first surgerized with either hM4Di DREADDs or eGFP null virus on a CaMKII promoter. Rats were allowed to recover for 1 week before testing on a stimulus-or action-based reversal learning task.* ***(C)*** *Initial learning of a rewarded stimulus, presented pseudorandomly on the left or right side of the touchscreen for eGFP (top), vlOFC hM4Di (middle), and BLA hM4Di (bottom).* ***(D)*** *Rats were always tested on a deterministic schedule before a probabilistic one. Shown are subsequent deterministic (100/0) and probabilistic (90/10) reversal "transitions," 3 sessions before and after each reversal. Drug order was counterbalanced such that on R2 and R4 animals received VEH if they were administered CNO first on R1 and R3, and vice versa. Chemogenetic inhibition of vlOFC abolishes changes in probability correct over the last 3 (pre) and first 3 (post)-reversal sessions, indicating impaired detection of reversal. In contrast, BLA inhibition had no impact on this detection. There was also no effect of CNO in eGFP group learning. **p<0.01 different than eGFP following ANOVA of pre-post difference.*

Action-based reversal learning. We also assessed accuracy (*probability correct*) around reversals for action-based reversal learning. As above, ANOVAs with virus and drug order as between subject factors were conducted on the mean accuracy change between one reversal and the next. Other than confirming the probabilistic reversal learning (R3) impairment for BLA

hM4Di (**Fig. 2**), there were no significant effects of virus groups on accuracy changes on any other reversal transition in action-based reversal learning (**Fig. 2-1).**

In summary, our results indicate that both BLA and vlOFC are required for learning action-based reversals in females. Conversely, vlOFC, but not BLA, is necessary for detecting stimulus-based, not action-based, reversals.

Win-Stay, Lose-Switch Strategies around reversals

Stimulus-based reversal learning. We also analyzed adaptive response strategies (*Win-Stay* and *Lose-Switch*) around reversals. ANOVAs with virus and drug order as between subject factors were conducted on mean Win-Stay or Lose-Switch between one reversal and the next. vlOFC hM4Di was significantly different from eGFP for *Win-Stay* R1-to-R2 [$F_{(1,24)}$=9.91, $p < 0.01$] and R2-to-R3 [$F_{(1,24)}$= 11.61, $p < 0.01$], but not R3-to-R4 [$F_{(1,24)}$= 1.49, $p = 0.24$]. Similarly, vlOFC

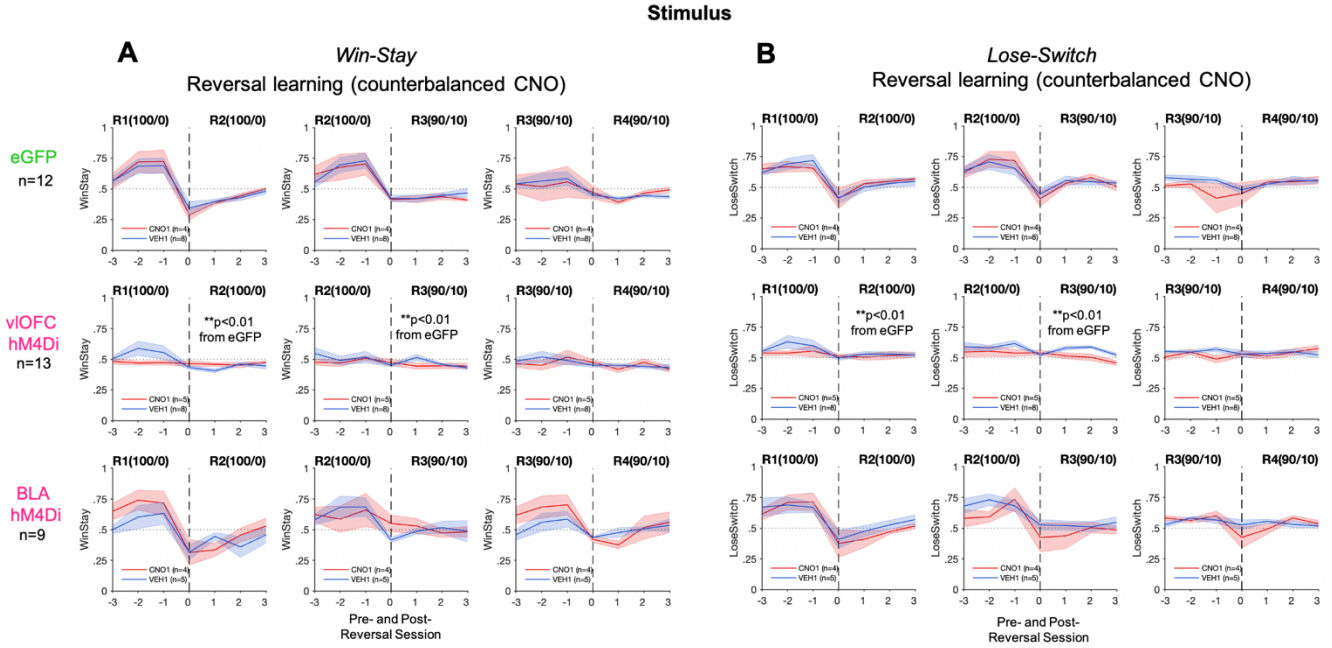

Figure 5. Inhibition of vlOFC, but not BLA, impairs the detection of deterministic stimulus-based reversals as measured by changes in Win-Stay, Lose-Switch strategies. (A) *Inhibition of vlOFC abolishes changes in Win-Stay over the last 3 (pre) and first 3 (post)-reversal sessions, indicating impaired detection of R1-R2, R2-R3. In contrast, BLA inhibition had no impact on this detection.* ***(B)*** *Inhibition of vlOFC abolishes changes in Lose-Shift over the last 3 (pre) and first 3 (post)-reversal sessions, indicating impaired detection of R1-R2, R2-R3. In contrast, BLA inhibition had no impact on this detection. There was also no effect of CNO in eGFP group learning. **p<0.01 different than eGFP following ANOVA of pre-post difference.*

hM4Di differed from eGFP for *Lose-Switch* in R1-to-R2 [$F_{(1,24)}$=6.00, p=0.023] and R2-to-R3 [$F_{(1,24)}$=5.00, p=0.036], but not R3-to-R4 [$F_{(1,24)}$=0.08, p=0.78], **Fig. 5**.

In contrast, BLA hM4Di was not significantly different from eGFP on changes in *Win-Stay* or *Lose-Switch* strategies around any of the reversals. Therefore, the results for these adaptive strategies reflect an identical pattern to that observed for *probability correct* for both vlOFC and BLA hM4Di, above.

Performance measures in reversal learning

We analyzed other performance measures including *initiation, choice,* and *reward latencies*, as proxies for attention, deliberation, and motivation, respectively (Harris, Aguirre et al., 2021; Ye et al., 2023). Given that the significant effects in accuracy were observed only in R1 and R3, we elected to analyze latency measures during those reversals only, with sex added as a covariate in the GLM models. In this case, a single measure per animal per reversal number was obtained: the GLM formula was as follows for R1 and R3: $\gamma \sim [1 + virus * drug + sex + (1 + drug| rat)]$.

Action-based reversal learning. vlOFC comparisons with eGFP controls did not yield any significant effects for R1; however, a significant drug effect emerged in R3, such that VEH-treated animals exhibited longer *initiation latencies* than when treated with CNO, irrespective of virus group (β_{virus} = -0.524, t(27) = -2.06, p = 0.049). Females tended to exhibit longer latencies than males, although this was only trending toward significance (β_{sex} = -0.353, t(27) = -1.96, p = 0.06).

When comparing BLA hM4Di vs. eGFP, we found a significant effect of virus in the full model (β_{virus} = 1.44, t(93) = 2.07, p = 0.039), with BLA hM4Di animals exhibiting longer initiation latencies than eGFP controls. We also observed a significant effect of virus for both *incorrect*

choice (R1: $\beta_{virus} = 0.38$, $t(27) = 4.68$ $p < 0.0001$; R3: $\beta_{virus} = 0.28$, $t(28) = 3.40$, $p = 0.002$) and *reward latencies* (R1: $\beta_{virus} = 0.34$, $t(30) = 3.78$ $p = 0.0007$; R3: $\beta_{virus} = 0.22$, $t(30) = 2.31$, $p = 0.03$), such that BLA hM4Di exhibited longer latencies compared to eGFP (**Fig. 2-2B**). This effect was not observed when comparing vlOFC hM4DI vs. eGFP (**Fig. 2-2A**).

Stimulus-based reversal learning. We similarly analyzed performance measures in stimulus-based reversal learning during R1 and R3, with sex added as a covariate to the model. We found a robust sex difference, with females committing more *initiation omissions* than males, irrespective of virus or drug, across reversals (vlOFC vs. eGFP: $\beta_{sex} = -88.52$, $t(80) = -2.40$ $p = 0.02$; BLA vs. eGFP: $\beta_{sex} = -80.38$, $t(89) = -2.14$, $p = 0.04$; **Fig. 4-2AB**). For *initiation latencies,* vlOFC hM4Di vs. eGFP analysis yielded an overall significant effect of sex as well ($\beta_{sex} = -3.23$, $t(81) = -3.72$, $p = 0.0004$), such that females exhibited longer initiation latencies than males across reversals (**Fig. 4-2CD**). BLA hM4Di vs. eGFP comparisons also yielded a significant effect of sex ($\beta_{sex} = -3.26$, $t(88) = -3.88$, $p = 0.0002$), however, there was also a significant interaction between reversal number, virus, drug, drug order, and sex ($\beta_{\text{reversal number x virus x drug x drug order x sex}} = 1.46$, $t(88) = 2.13$, $p = 0.036$), which justified analysis of each reversal number separately. Subsequent analyses revealed a significant effect of sex for both R1($\beta_{sex} = -1.58$, $t(24) = -3.40$, $p = 0.002$), and R3 ($\beta_{sex} = -0.67$, $t(24) = -2.42$, $p = 0.024$), with females exhibiting longer initiation latencies than males, irrespective of virus or drug (**Fig. 4-2D**).

In summary, across all reversals BLA hM4Di animals exhibited slower deliberation speed during incorrect choices, and took longer to collect reward compared to eGFP controls in the action-based task. In contrast, sex emerged as a strong predictor of performance measures, but not learning, in the stimulus-based task, as it accounted for much of the variance in initiation omissions and latencies.

Discussion

We used a chemogenetic approach to transiently inactivate neurons in either vlOFC or BLA to assess how these regions are involved in different aspects of reversal learning. Although the role of OFC in reversal learning has been instantiated in different paradigms using visual stimuli and cues (Izquierdo et al., 2013; Piantadosi et al., 2018; Hervig et al., 2020; Alsio et al., 2021) as well as olfactory ones (Schoenbaum et al., 2003; Kim and Ragozzino, 2005), several groups also report a strong role for OFC in action (spatial)-based reversal learning (Dalton et al., 2016; Groman et al., 2019; Verharen et al., 2020). Almost all of these reversal learning investigations have involved irreversible lesions or baclofen/muscimol inactivations of OFC. Testing both types with a chemogenetic approach, here we found that different aspects of both action- and stimulus-based reversal learning rely on vlOFC.

In parallel, the specific role of BLA in stimulus- vs. action-based reversal learning is poorly understood given mixed results (Schoenbaum et al., 2003; Izquierdo and Murray, 2004; Churchwell et al., 2009; Hervig et al., 2020). Recent studies suggest amygdala may be involved in both types of learning (Taswell et al., 2021; Keefer and Petrovich, 2022) as BLA activity is modulated by violations in reward expectations generally, which are not association-specific (Roesch et al., 2012). To probe this, we tested animals on both stimulus- and action-based tasks and found that BLA is more selectively involved in action-based reversal learning.

As additional motivations for the present study, several reports suggest that neural recruitment in reversal learning may depend on certainty of rewards (Boulougouris et al., 2007; Boulougouris and Robbins, 2009; Ward et al., 2015; Costa et al., 2016; Dalton et al., 2016; Piantadosi et al., 2018; Verharen et al., 2020). To further understand this, we tested animals on both deterministic (100/0) and probabilistic reversals (90/10). We found vlOFC to be involved in

the detection of stimulus-based reversals and initial learning of both deterministic and probabilistic learning across stimuli and actions, whereas BLA was more selectively involved in probabilistic reversal learning of actions.

Finally, due to the sparsity of research probing sex differences in flexible learning and decision making (Orsini and Setlow, 2017; Grissom and Reyes, 2019; Orsini et al., 2022; Cox et al., 2023), where an overwhelming number of reversal learning studies include only males (Schoenbaum et al., 2003; Izquierdo et al., 2013; Dalton et al., 2016; Groman et al., 2019; Hervig et al., 2020; Verharen et al., 2020), we included both male and female rats here. We found sex-dependent contributions of both vlOFC and BLA in action-based reversal learning. We elaborate on these findings within the context of the existing literature below.

Similar recruitment of BLA and vlOFC during action-based reversal learning

All animals learned to flexibly adjust their responses following deterministic and probabilistic reversals, indicating successful remapping of reward contingencies as accuracy increased across trials. Importantly, we found no effect of CNO in eGFP animals, suggesting that it was activation of hM4Di receptors in BLA and vlOFC that were crucial to any impairments observed. After considering sex as a covariate in our analyses, we determined that vlOFC was necessary for learning of first deterministic (R1) and probabilistic reversal (R3). This is consistent with findings following pharmacological inactivations or lesions of OFC (Boulougouris et al., 2007; Boulougouris and Robbins, 2009; Dalton et al., 2016; Piantadosi et al., 2018; Verharen et al., 2020).

BLA inhibition was not expected to impair deterministic reversal learning as it is thought to be mostly recruited when there is some level of uncertainty, e.g., probabilistic outcomes (Roesch

et al., 2012). Amygdala-lesioned monkeys are also impaired on action(spatial)-based probabilistic reversal learning, exhibiting decreased probability of choosing the better option, and increased switching behavior following negative outcomes (Taswell et al., 2021). BLA may indeed be critical in generating prediction error signals following changes in reward associations (Esber et al., 2012; Roesch et al., 2012; Iordanova et al., 2021), with particular involvement in detecting unexpected upshifts or downshifts in value (Roesch et al., 2010; Stolyarova and Izquierdo, 2017). Our finding of attenuated learning of probabilistic reversal R3 suggests it is the misleading feedback that most engages BLA. Future investigations should probe the role of BLA in initial learning of probabilistic outcomes, without reversal experience.

vlOFC, but not BLA, is necessary for detection of reversals in stimulus-based learning

As described above, unlike the ease of action-based reversal learning, rats exhibited difficulty learning reversals of stimulus-reward contingencies, as previously reported (Harris, Aguirre et al., 2021). Thus, instead of examining acquisition curves which reached asymptote slightly above chance, we elected to study detection and adjustment to reversals by comparing accuracy and strategy prior to and after a reversal occurred. Furthermore, this enabled assessment about whether prior inhibition affected future detection of reversals and whether this varied by transition type [i.e., between deterministic reversals (R1→R2), deterministic and probabilistic reversal (R2→R3), or probabilistic reversals (R3→R4)]. We found that vlOFC, but not BLA, inhibition produced a failure in detecting first deterministic and first probabilistic reversal. This pattern was not observed in animals that received VEH during the first deterministic reversal, suggesting vlOFC needs to be "online" when experiencing a reversal for the very first time as this determines how flexibly animals respond to future reversals. Employment of adaptive strategies

741 matched this effect, such that vlOFC-inhibited animals did not employ either win-stay or lose-
742 switch strategies after the first reversal. That vlOFC inhibition did not impair the ability to detect
743 transitions between probabilistic reversals (R3→R4) supports the idea that other brain regions may
744 be recruited when the probabilistic reward contingencies have already been established. The role
745 of OFC in establishing an "expected uncertainty" (Soltani and Izquierdo, 2019) has been
746 instantiated experimentally in several recent studies using different methodologies (Namboodiri et
747 al., 2019; Namboodiri et al., 2021; Jenni et al., 2022), and we add detection and adjustment to
748 stimulus-based reversals to this evidence. In contrast, BLA inhibition did not result in any
749 impairment in the ability to detect and flexibly adjust to reversals, regardless of whether they were
750 deterministic or probabilistic.
751
752 ***Females exhibited poorer action-based reversal learning following vlOFC and BLA inhibition***
753 We found a significant sex-dependent effect of vlOFC inhibition for both the first
754 deterministic (R1) and the first probabilistic reversal (R3), and a similar effect in females following
755 BLA inhibition for R3. Importantly, although BLA hM4Di males also exhibited the general pattern
756 of attenuated learning of R3, the effect of BLA inhibition was largely driven by females. These
757 findings were unexpected as we did not anticipate sex differences in the recruitment of brain
758 regions involved in action-based reversal learning, mostly due to the lack of studies to date
759 investigating the effect of sex. Nonetheless, there is evidence that OFC is differentially activated
760 in males and females during risky decisions such that activity in the lateral OFC is inversely
761 correlated with the proportion of advantageous choices in female, but not in male, rats (van Hasselt
762 et al., 2012), with similar effects found in humans (Bolla et al., 2004) and in amygdala (Dreher et
763 al., 2007).

764 Unfortunately, the potential of estrous-driven behavioral variation in females has
765 commonly been used as a rationale for excluding female rodents in behavioral neuroscience
766 research (Beery and Zucker, 2011; McCarthy et al., 2012; Shansky and Murphy, 2021). However,
767 our findings, and other recent studies of cortical circuits exhibiting sex-mediated influences on
768 reward-motivated behavior (Cox et al., 2023) should instead encourage the inclusion of both sexes
769 in experiments as clear patterns emerge. Additionally, the learning impairment observed in females
770 following vlOFC and BLA inhibition may not be primarily due to fluctuations in hormone levels,
771 but rather may reflect differential adoption of strategies between sexes (Grissom and Reyes, 2019;
772 Chen et al., 2021a; Chen et al., 2021b). Consistent with this view, our results based on estimated
773 RL model parameters suggests differential mechanisms for adjustment to reversals between males
774 and females. Specifically, we found some evidence for differential effects on γ_d, where the decay
775 rate for unchosen action values was greater for females than males. This is consistent with a
776 previous study which also observed similar disruption in retention of action values after ablating
777 OFC neurons projecting to BLA (Groman et al., 2019). Given that study involved only male rats
778 and involved a stronger manipulation than our chemogenetic approach (i.e., one that caused
779 pathway-specific neuronal apoptosis), it is plausible their observed effect on γ_d would have been
780 different between females and males.
781
782 ***Stimulus-based vs. action-based learning***
783 Interestingly, we discovered task order to be significant in rats' ability to learn to
784 discriminate stimuli: stimulus→action was learned much more readily than action→stimulus. This
785 can be explained by noting that rats are heavily biased to acquire spatial associations (Wright et
786 al., 2019), and reinforcing this already-strong learning likely inhibits the ability to learn

787 associations where spatial information should be ignored. In contrast, nonhuman primates are able
788 to quickly transition between "what" vs. "where" blocks of trials (Rothenhoefer et al., 2017;
789 Taswell et al., 2021). Nonetheless, learning both types of associations is crucial for flexibility
790 required in naturalistic environments and thus, it is important to examine how stimulus-based and
791 action-based learning systems interact with each other (Soltani & Koechlin, 2022). Moreover,
792 although the role of OFC in stimulus- or cue-based reversal learning has been probed using
793 olfactory and visual stimuli, more viral-mediated approaches employing targeted chemogenetic
794 and optogenetic manipulations across sensory modalities in both males and females are warranted.
795
796 ***Conclusion***
797 The present results suggest similar roles for vlOFC and BLA in flexible action-based
798 learning, but a more specialized role for vlOFC in setting adjustments in stimulus-based learning.
799 Additionally, our findings underscore the importance of including both male and female animals
800 in behavioral neuroscience studies, adding to the mounting evidence of sex-modulated learning
801 flexibility (Chen et al., 2021a; Chen et al., 2021b).
802

References

Alsio J, Lehmann O, McKenzie C, Theobald DE, Searle L, Xia J, Dalley JW, Robbins TW (2021) Serotonergic Innervations of the Orbitofrontal and Medial-prefrontal Cortices are Differentially Involved in Visual Discrimination and Reversal Learning in Rats. Cereb Cortex 31:1090-1105.

Asrican B, Song J (2021) Extracting meaningful circuit-based calcium dynamics in astrocytes and neurons from adult mouse brain slices using single-photon GCaMP imaging. STAR Protoc 2:100306.

Barreiros IV, Panayi MC, Walton ME (2021a) Organization of Afferents along the Anterior-posterior and Medial-lateral Axes of the Rat Orbitofrontal Cortex. Neuroscience 460:53-68.

Barreiros IV, Ishii H, Walton ME, Panayi MC (2021b) Defining an orbitofrontal compass: Functional and anatomical heterogeneity across anterior-posterior and medial-lateral axes. Behav Neurosci 135:165-173.

Beery AK, Zucker I (2011) Sex bias in neuroscience and biomedical research. Neurosci Biobehav Rev 35:565-572.

Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214-1221.

Bolla KI, Eldreth DA, Matochik JA, Cadet JL (2004) Sex-related Differences in a Gambling Task and Its Neurological Correlates. Cerebral Cortex 14:1226-1232.

Boulougouris V, Robbins TW (2009) Pre-surgical training ameliorates orbitofrontal-mediated impairments in spatial reversal learning. Behavioural Brain Research 197:469-475.

Boulougouris V, Dalley JW, Robbins TW (2007) Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. Behavioural Brain Research 179:219-228.

Chen CS, Knep E, Han A, Ebitz RB, Grissom NM (2021a) Sex differences in learning from exploration. Elife 10.

Chen CS, Ebitz RB, Bindas SR, Redish AD, Hayden BY, Grissom NM (2021b) Divergent Strategies for Learning in Males and Females. Curr Biol 31:39-50 e34.

Churchwell JC, Morris AM, Heurtelou NM, Kesner RP (2009) Interactions between the prefrontal cortex and amygdala during delay discounting and reversal. Behav Neurosci 123:1185-1196.

Corbit LH, Balleine BW (2005) Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. J Neurosci 25:962-970.

Costa KM, Scholz R, Lloyd K, Moreno-Castilla P, Gardner MPH, Dayan P, Schoenbaum G (2023) The role of the lateral orbitofrontal cortex in creating cognitive maps. Nat Neurosci 26:107-115.

Costa VD, Dal Monte O, Lucas DR, Murray EA, Averbeck BB (2016) Amygdala and Ventral Striatum Make Distinct Contributions to Reinforcement Learning. Neuron 92:505-517.

Cox J, Minerva AR, Fleming WT, Zimmerman CA, Hayes C, Zorowitz S, Bandi A, Ornelas S, McMannon B, Parker NF, Witten IB (2023) A neural substrate of sex-dependent modulation of motivation. Nature Neuroscience 26:274-284.

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. J Neurosci 36:1996-2006.

Dreher J-C, Schmidt PJ, Kohn P, Furman D, Rubinow D, Berman KF (2007) Menstrual cycle phase modulates reward-related neural function in women. Proc Natl Acad Sci U S A 104:2465-2470.

Esber GR, Roesch MR, Bali S, Trageser J, Bissonette GB, Puche AC, Holland PC, Schoenbaum G (2012) Attention-related Pearce-Kaye-Hall signals in basolateral amygdala require the midbrain dopaminergic system. Biol Psychiatry 72:1012-1019.

Ghasemi A, Jeddi S, Kashfi K (2021) The laboratory rat: Age and body weight matter. EXCLI J 20:1431-1445.

Grissom NM, Reyes TM (2019) Let's call the whole thing off: evaluating gender and sex differences in executive function. Neuropsychopharmacology 44:86-96.

Groman SM, Keistler C, Keip AJ, Hammarlund E, DiLeone RJ, Pittenger C, Lee D, Taylor JR (2019) Orbitofrontal Circuits Control Multiple Reinforcement-Learning Processes. Neuron 103:734-746 e733.

Harris C, Aguirre C, Kolli S, Das K, Izquierdo A, Soltani A (2021) Unique features of stimulus-based probabilistic reversal learning. Behav Neurosci 135:550-570.

Hart EE, Blair GJ, O'Dell TJ, Blair HT, Izquierdo A (2020) Chemogenetic Modulation and Single-Photon Calcium Imaging in Anterior Cingulate Cortex Reveal a Mechanism for Effort-Based Decisions. J Neurosci 40:5628-5643.

Hervig ME, Fiddian L, Piilgaard L, Bozic T, Blanco-Pozo M, Knudsen C, Olesen SF, Alsio J, Robbins TW (2020) Dissociable and Paradoxical Roles of Rat Medial and Lateral Orbitofrontal Cortex in Visual Serial Reversal Learning. Cereb Cortex 30:1016-1029.

Iordanova MD, Yau JO-Y, McDannald MA, Corbit LH (2021) Neural substrates of appetitive and aversive prediction error. Neurosci Biobehav Rev 123:337-351.

Izquierdo A (2017) Functional Heterogeneity within Rat Orbitofrontal Cortex in Reward Learning and Decision Making. J Neurosci 37:10529-10540.

Izquierdo A, Murray EA (2004) Combined Unilateral Lesions of the Amygdala and Orbital Prefrontal Cortex Impair Affective Processing in Rhesus Monkeys. Journal of Neurophysiology 91:2023-2039.

Izquierdo A, Darling C, Manos N, Pozos H, Kim C, Ostrander S, Cazares V, Stepp H, Rudebeck PH (2013) Basolateral amygdala lesions facilitate reward choices after negative feedback in rats. J Neurosci 33:4105-4109.

Janak PH, Tye KM (2015) From circuits to behaviour in the amygdala. Nature 517:284-292.

Jang AI, Costa VD, Rudebeck PH, Chudasama Y, Murray EA, Averbeck BB (2015) The Role of Frontal Cortical and Medial-Temporal Lobe Brain Areas in Learning a Bayesian Prior Belief on Reversals. J Neurosci 35:11751-11760.

Jenni NL, Rutledge G, Floresco SB (2022) Distinct Medial Orbitofrontal-Striatal Circuits Support Dissociable Component Processes of Risk/Reward Decision-Making. J Neurosci 42:2743-2755.

Keefer SE, Petrovich GD (2022) Necessity and recruitment of cue-specific neuronal ensembles within the basolateral amygdala during appetitive reversal learning. Neurobiol Learn Mem 194:107663.

Kim J, Ragozzino ME (2005) The involvement of the orbitofrontal cortex in learning under changing task contingencies. Neurobiol Learn Mem 83:125-133.

Lichtenberg NT, Pennington ZT, Holley SM, Greenfield VY, Cepeda C, Levine MS, Wassum KM (2017) Basolateral Amygdala to Orbitofrontal Cortex Projections Enable Cue-Triggered Reward Expectations. J Neurosci 37:8374-8384.

Malvaez M, Shieh C, Murphy MD, Greenfield VY, Wassum KM (2019) Distinct cortical-amygdala projections drive reward value encoding and retrieval. Nat Neurosci 22:762-769.

McCarthy MM, Arnold AP, Ball GF, Blaustein JD, De Vries GJ (2012) Sex differences in the brain: the not so inconvenient truth. J Neurosci 32:2241-2247.

Namboodiri VMK, Hobbs T, Trujillo-Pisanty I, Simon RC, Gray MM, Stuber GD (2021) Relative salience signaling within a thalamo-orbitofrontal circuit governs learning rate. Curr Biol 31:5176-5191.e5175.

Namboodiri VMK, Otis JM, van Heeswijk K, Voets ES, Alghorazi RA, Rodriguez-Romaguera J, Mihalas S, Stuber GD (2019) Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. Nat Neurosci 22:1110-1121.

Orsini CA, Setlow B (2017) Sex differences in animal models of decision making. J Neurosci Res 95:260-269.

Orsini CA, Truckenbrod LM, Wheeler A-R (2022) Regulation of sex differences in risk-based decision making by gonadal hormones: Insights from rodent models. Behav Processes 200:104663-104663.

Pachitariu M, Stringer C, Dipoppa M, Schröder S, Rossi LF, Dalgleish H, Carandini M, Harris KD (2017) Suite2p: beyond 10,000 neurons with standard two-photon microscopy. bioRxiv:061507.

Piantadosi PT, Lieberman AG, Pickens CL, Bergstrom HC, Holmes A (2018) A novel multichoice touchscreen paradigm for assessing cognitive flexibility in mice. Learn Mem 26:24-30.

Roesch MR, Calu DJ, Esber GR, Schoenbaum G (2010) Neural correlates of variations in event processing during learning in basolateral amygdala. J Neurosci 30:2464-2471.

Roesch MR, Esber GR, Bryden DW, Cerri DH, Haney ZR, Schoenbaum G (2012) Normal aging alters learning and attention-related teaching signals in basolateral amygdala. J Neurosci 32:13137-13144.

Rothenhoefer KM, Costa VD, Bartolo R, Vicario-Feliciano R, Murray EA, Averbeck BB (2017) Effects of ventral striatum lesions on stimulus-based versus action-based reinforcement learning. J Neurosci 37: 6902-6914.

Rudebeck PH, Murray EA (2008) Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. J Neurosci 28:8338-8343.

Schoenbaum G, Setlow B, Nugent SL, Saddoris MP, Gallagher M (2003) Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. Learn Mem 10:129-140.

Shansky RM, Murphy AZ (2021) Considering sex as a biological variable will require a global shift in science culture. Nature Neuroscience 24:457-464.

Sias AC, Morse AK, Wang S, Greenfield VY, Goodpaster CM, Wrenn TM, Wikenheiser AM, Holley SM, Cepeda C, Levine MS, Wassum KM (2021) A bidirectional corticoamygdala circuit for the encoding and retrieval of detailed reward memories. Elife 10.

Soltani A, Izquierdo A (2019) Adaptive learning under expected and unexpected uncertainty. Nat Rev Neurosci 20:635-644.

Soltani A, Koechlin E (2022) Computational models of adaptive behavior and prefrontal cortex. Neuropsychopharmacology 47: 58-71.

940 Stalnaker TA, Roesch MR, Calu DJ, Burke KA, Singh T, Schoenbaum G (2007) Neural
941 correlates of inflexible behavior in the orbitofrontal-amygdalar circuit after cocaine
942 exposure. Ann N Y Acad Sci 1121:598-609.
943 Stolyarova A, Izquierdo A (2017) Complementary contributions of basolateral amygdala and
944 orbitofrontal cortex to value learning under uncertainty. Elife 6.
945 Stolyarova A, Rakhshan M, Hart EE, O'Dell TJ, Peters MAK, Lau H, Soltani A, Izquierdo A
946 (2019) Contributions of anterior cingulate cortex and basolateral amygdala to decision
947 confidence and learning under uncertainty. Nat Commun 10:4704.
948 Tada M, Takeuchi A, Hashizume M, Kitamura K, Kano M (2014) A highly sensitive fluorescent
949 indicator dye for calcium imaging of neural activity in vitro and in vivo. Eur J Neurosci
950 39:1720-1728.
951 Taswell CA, Costa VD, Basile BM, Pujara MS, Jones B, Manem N, Murray EA, Averbeck BB
952 (2021) Effects of Amygdala Lesions on Object-Based Versus Action-Based Learning in
953 Macaques. Cereb Cortex 31:529-546.
954 Tye KM, Janak PH (2007) Amygdala neurons differentially encode motivation and
955 reinforcement. J Neurosci 27:3937-3945.
956 van Hasselt FN, de Visser L, Tieskens JM, Cornelisse S, Baars AM, Lavrijsen M, Krugers HJ,
957 van den Bos R, Joëls M (2012) Individual variations in maternal care early in life
958 correlate with later life decision-making and c-fos expression in prefrontal subregions of
959 rats. PLoS One 7:e37820-e37820.
960 Verharen JPH, den Ouden HEM, Adan RAH, Vanderschuren L (2020) Modulation of value-
961 based decision making behavior by subregions of the rat prefrontal cortex.
962 Psychopharmacology (Berl) 237:1267-1280.
963 Ward RD, Winiger V, Kandel ER, Balsam PD, Simpson EH (2015) Orbitofrontal cortex
964 mediates the differential impact of signaled-reward probability on discrimination
965 accuracy. Front Neurosci 9:230.
966 Wassum KM, Izquierdo A (2015) The basolateral amygdala in reward learning and addiction.
967 Neurosci Biobehav Rev 57:271-283.
968 Winstanley CA, Floresco SB (2016) Deciphering Decision Making: Variation in Animal Models
969 of Effort- and Uncertainty-Based Choice Reveals Distinct Neural Circuitries Underlying
970 Core Cognitive Processes. J Neurosci 36:12069-12079.
971 Wright SL, Martin GM, Thorpe CM, Haley K, Skinner DM (2019) Distance and direction, but
972 not light cues, support response reversal learning. Learning & Behavior 47:38-46.
973 Ye T, Romero-Sosa JL, Rickard A, Aguirre CG, Wikenheiser AM, Blair HT, Izquierdo A (2023)
974 Theta oscillations in anterior cingulate cortex and orbitofrontal cortex differentially
975 modulate accuracy and speed in flexible reward learning. Oxford Open Neuroscience
976 kvad005.
977 Zimmermann KS, Li CC, Rainnie DG, Ressler KJ, Gourley SL (2018) Memory Retention
978 Involves the Ventrolateral Orbitofrontal Cortex: Comparison with the Basolateral
979 Amygdala. Neuropsychopharmacology 43:373-383.
980

981 **Extended Data**

982

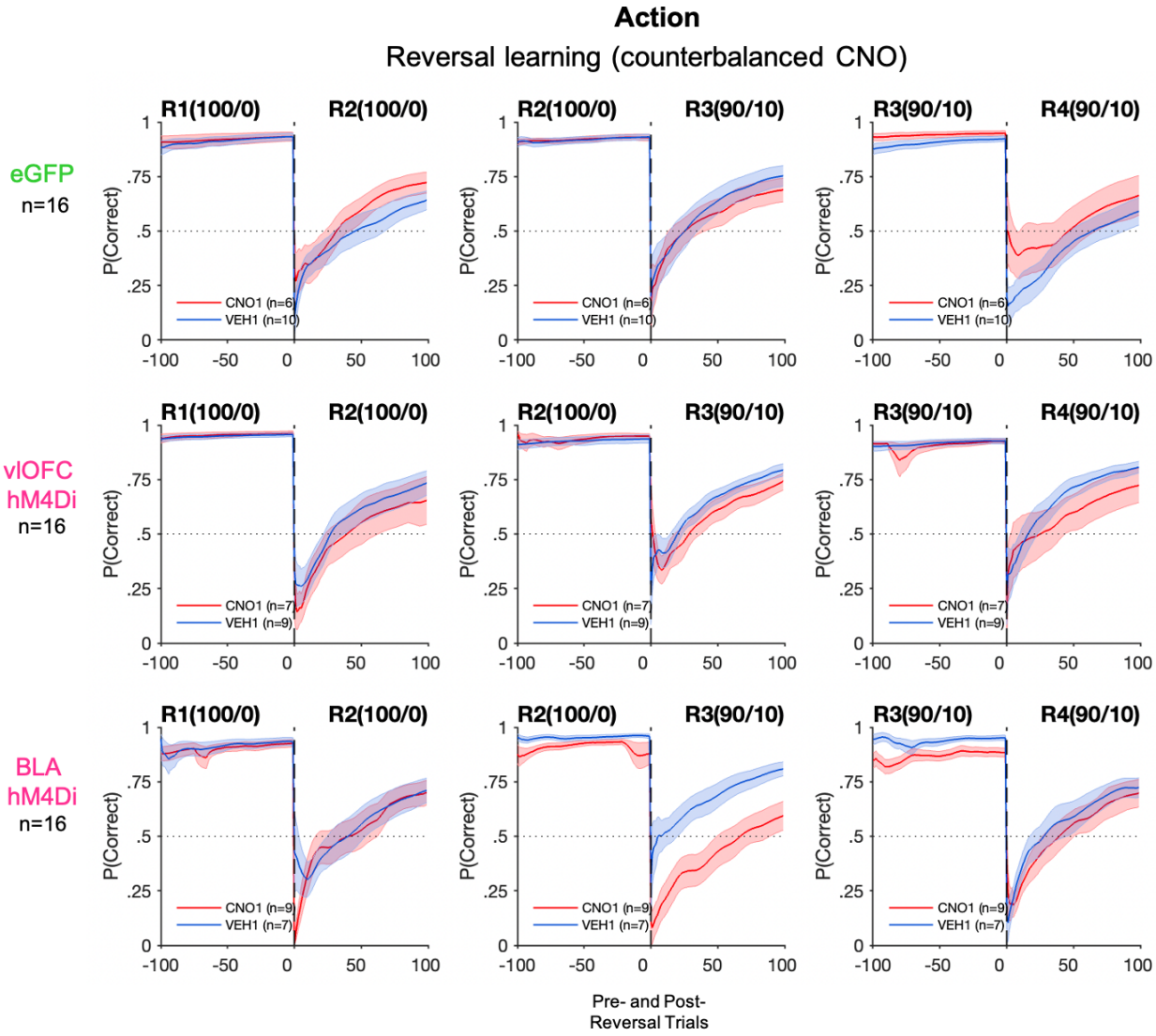

Figure 2-1. Neither BLA or vlOFC inhibition affected the detection of action-based reversals as measured by probability correct adjustments. *Rats were always tested on a deterministic schedule before a probabilistic one. Shown are deterministic (100/0) and probabilistic (90/10) reversal "transitions," 100 trials before and after each reversal. Drug order was counterbalanced such that on R2 and R4 animals received VEH if they were administered CNO first on R1 and R3, and vice versa. Inhibition of BLA impaired probabilistic reversal learning as indicated by selectively poor performance in 90/10 at the beginning and end of that reversal. OFC inhibition resulted in no impairment around these reversals. There was also no effect of CNO in the eGFP group.*

983

984

985

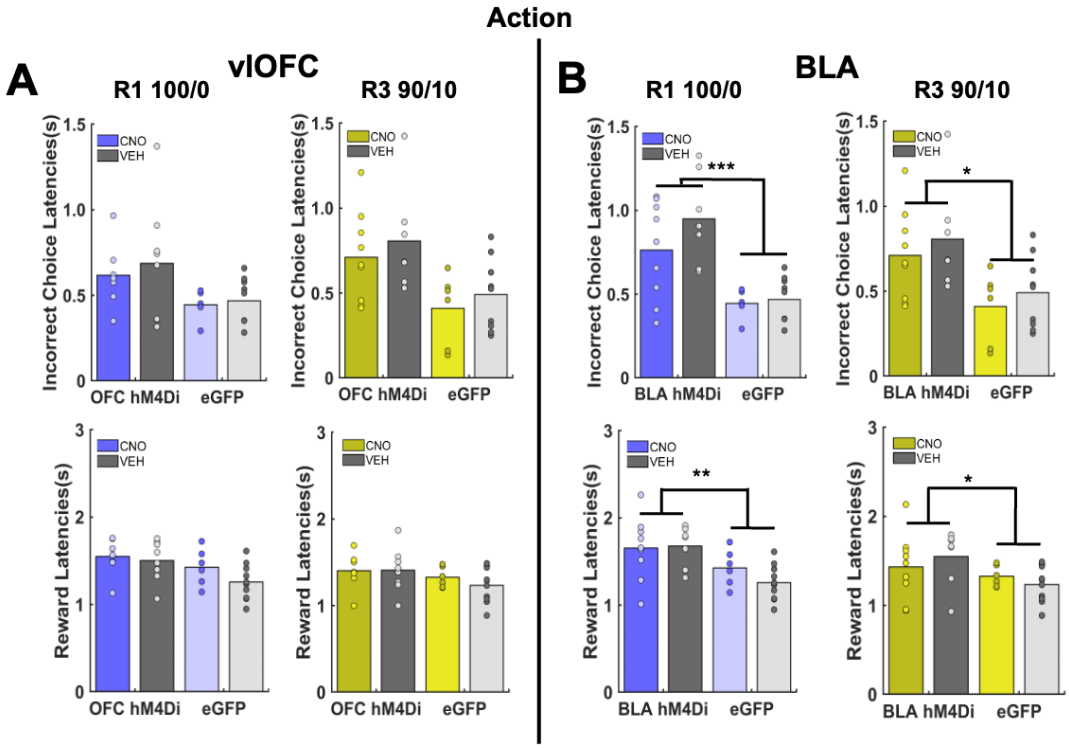

Figure 2-2. BLA, but not vlOFC, hM4Di exhibited longer incorrect choice and reward latencies than eGFP during action-based reversals. (A) *There were no significant differences in incorrect choice and reward latencies when comparing vlOFC hM4Di vs. eGFP controls for any reversals.* ***(B)*** *BLA hM4Di animals exhibited longer incorrect choice and reward latencies than eGFP controls during reversals. ***p<0.0001, **p<0.001, *p<0.05*

986

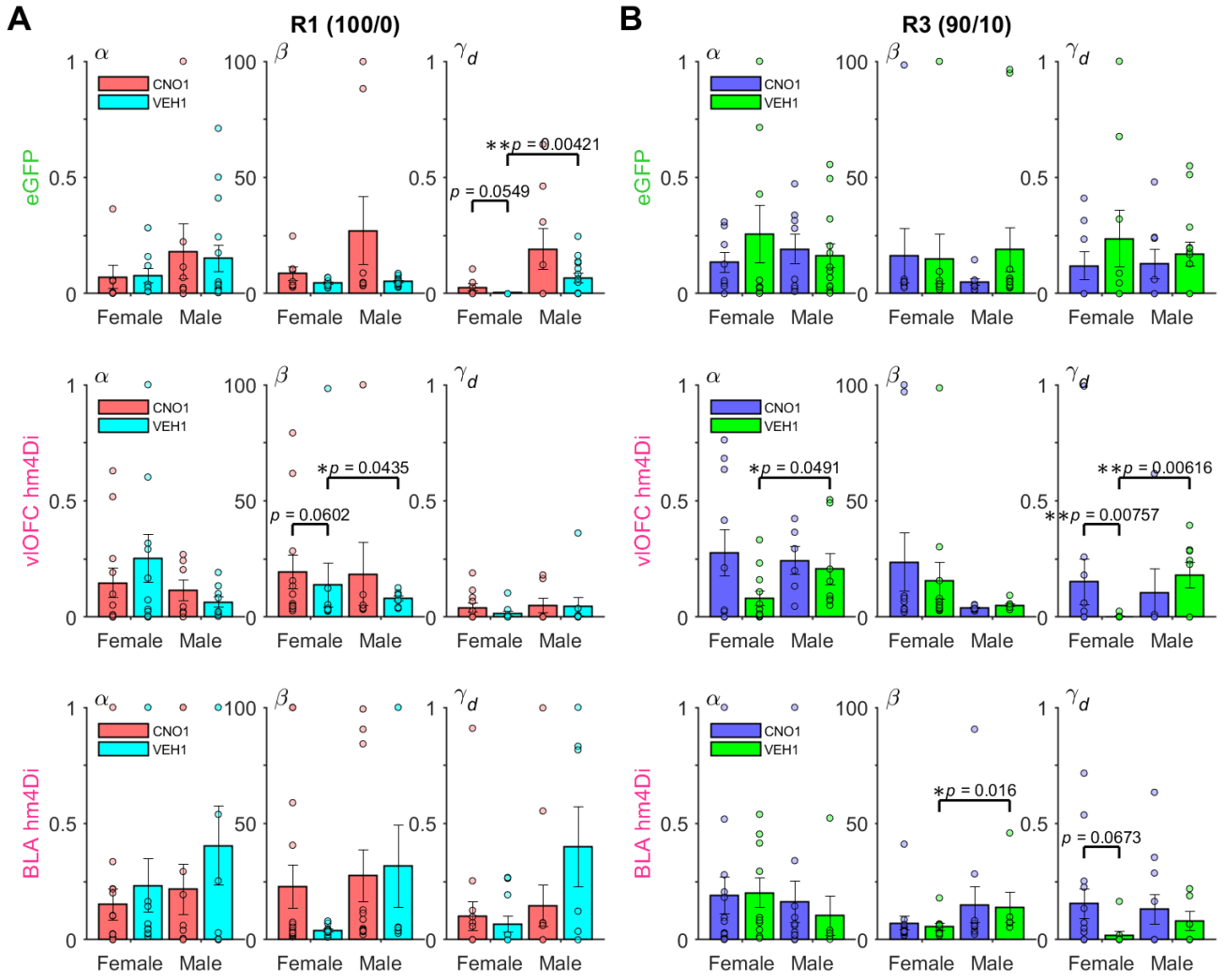

Figure 3-1. Reinforcement learning model fit to choice behavior indicates differential effect on the decay parameter between male and female rats during deterministic and probabilistic action-based reversals. *Plotted are estimated model parameters for $RL1_{decay}$ fitted to each session during deterministic (100/0)* ***(A)*** *and probabilistic (90/10)* ***(B)*** *reversals. α = learning rate, β = inverse temperature, γ_d = decay rate for unchosen option. Drug order was counterbalanced such that on R2 and R4 (not shown) animals received VEH if they were administered CNO first on R1 and R3, and vice versa. Female eGFP group exhibited overall lower γ_d than male eGFP rats during the first deterministic reversal R1. Chemogenetic inhibition of vlOFC or BLA increased γ_d on the first probabilistic reversal R3 for female rats. Only comparisons with p<0.075 are shown. **p<0.01, *p<0.05.*

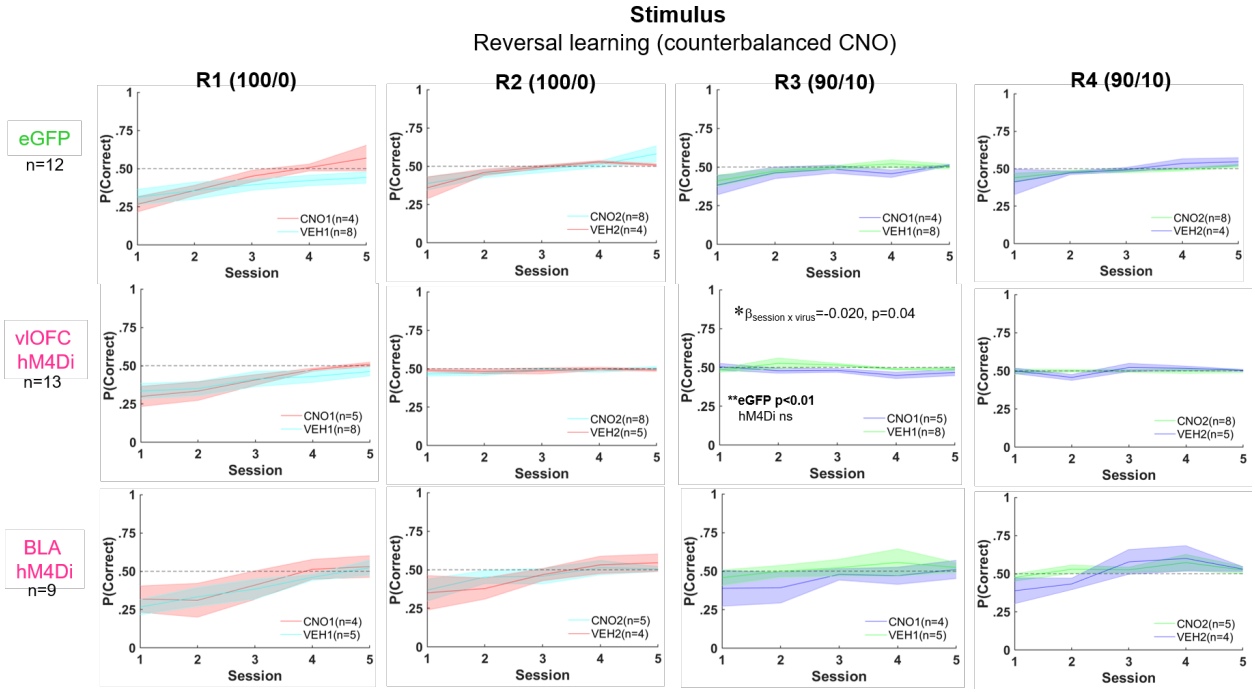

Figure 4-1. Probability correct during stimulus-based reversals. *Cumulative P(correct) for first 5 sessions of each deterministic (100/0) and probabilistic (90/10) reversal. Drug order was counterbalanced such that on R2 and R4 animals received VEH if they were administered CNO first on R1 and R3, and vice versa. Despite most animals reaching criterion on initial learning (Fig. 2C), animals exhibited poor reversal learning. Performance around reversals (3 sessions before and after each reversal) is shown in Fig.2. Reversal learning was particularly flat for R3 (first probabilistic reversal) for vlOFC hM4Di group compared to eGFP. Bonferroni-corrected post-hocs following mixed-effects GLM with sex as a covariate fixed factor wherein a session x virus interaction was found resulted in **p<0.01 effect of session in eGFP, not in hM4Di.*

994

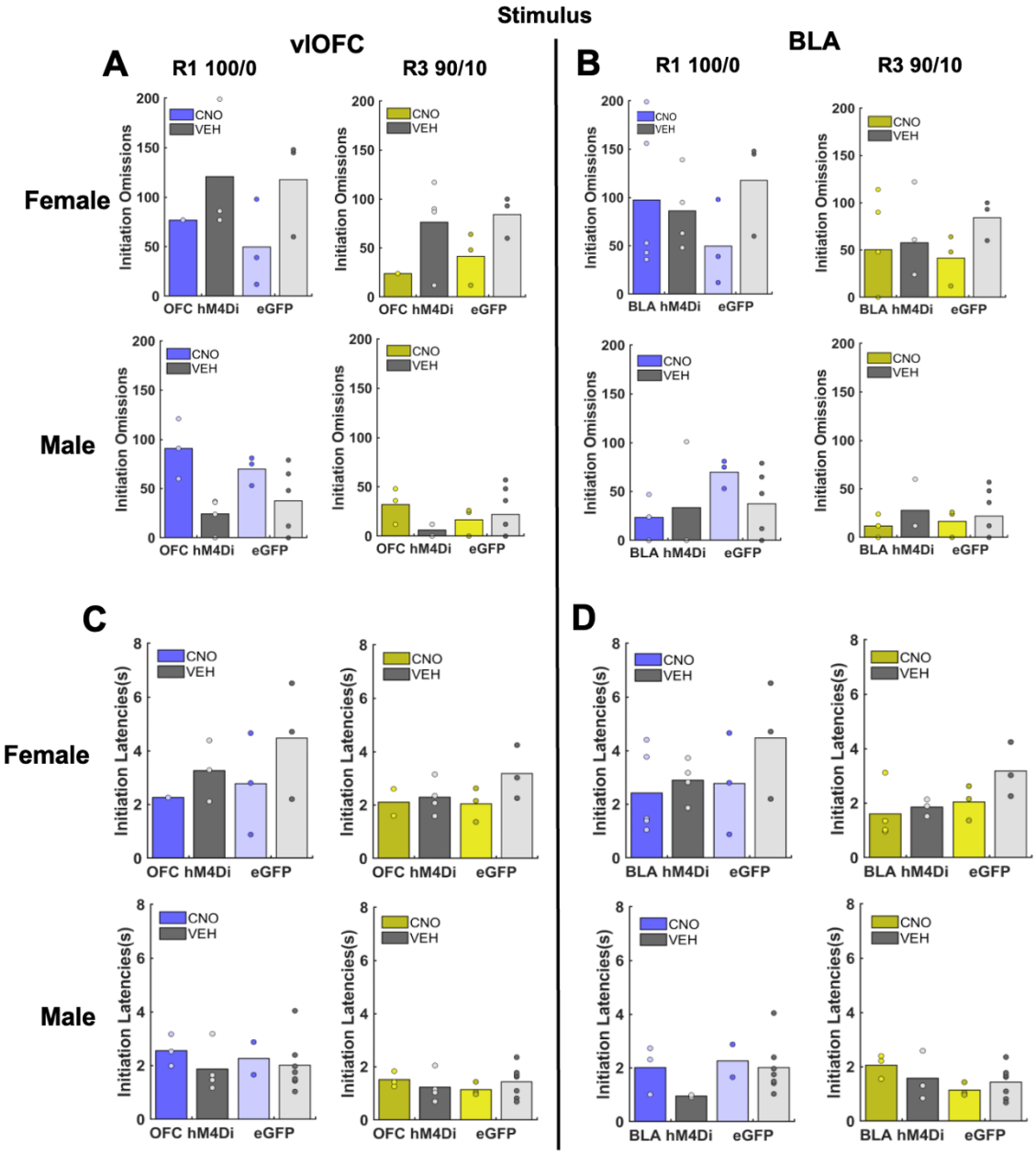

Figure 4-2. Females commit more initiation omissions and take longer to initiate trials than males in stimulus-based reversals. (A-B) *Females committed more initiation omissions than males irrespective of virus group or drug condition across reversals.* ***(C-D)*** *Females take longer to initiate trials than males irrespective of virus group or drug condition across reversals.*

995