Theta oscillations in frontal cortex differentially modulate accuracy and speed in flexible stimulus- and action-based learning

Running Title: *Frontal cortex theta in flexible learning*

Tony Ye[1], Andrew M. Wikenheiser[1-4], Hugh T. Blair[1-3], and Alicia Izquierdo*[1-4]

[1] Department of Psychology, UCLA, Los Angeles, CA 90095, USA
[2] The Brain Research Institute, UCLA, Los Angeles, CA 90095, USA
[3] Integrative Center for Learning and Memory, UCLA, Los Angeles, CA 90095, USA
[4] Integrative Center for Addictions, UCLA, Los Angeles, CA 90095, USA

Author contributions: TY and AI designed the research; TY performed the research; TY and AI analyzed the data; TY, AMW, HTB, and AI interpreted the data; HTB and AI acquired funding for the project; TY and AI wrote the paper.

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

Manuscript Information:

Pages 40
Figures 7 Supplementary Figures 7

Words: Abstract 248
Intro 742
Discussion 1280

ABSTRACT

Flexible reward learning relies on frontal cortex, with substantial evidence indicating the anterior cingulate cortex (ACC) and orbitofrontal cortex (OFC) play important roles. Recent studies in both rat and macaque suggest theta oscillations (5-10 Hz) may be a spectral signature that coordinates this learning. However, causal network-level interactions between ACC and OFC in flexible learning remain unclear. We investigated the learning of stimulus-outcome (S-O) and action-outcome (A-O) associations using a combination of simultaneous *in-vivo* electrophysiology in dorsal ACC and ventral OFC, partnered with bilateral inhibitory DREADDs in ACC. In freely-behaving male and female rats and using a within-subject design, we examined accuracy and speed of response across distinct and precisely-defined trial epochs including correct choice, incorrect choice, and reward collection. We report significant modulation of accuracy by theta power in both ACC and OFC. Both ACC and OFC theta oscillations consistently signaled accuracy in the initial discrimination phase, whereas it was OFC theta alone in the reversal phases of both S-O and A-O learning. Theta power in either region was not correlated with deliberation speed. Indeed, theta modulation of accuracy was dissociable from its involvement in speed of response which was affected by ACC inhibition, most prominently in A-O learning. Results are discussed with reference to the nonhuman primate literature, where there is by contrast more reported specialization of OFC and ACC in flexible learning of stimuli and actions. The present results also point to a specific role of OFC theta in signaling a reversal of either kind.

Keywords: **reversal learning, oscillations, prefrontal cortex, theta, DREADDs**

INTRODUCTION

Both humans and nonhumans alike make a multitude of choices based on adaptations to their environment in order to maximize rewards while minimizing losses. This type of flexible reward learning is thought to rely on the prefrontal cortex. Substantial evidence in primates indicates that anterior cingulate cortex (ACC) and orbitofrontal cortex (OFC) have essential roles in learning about actions and stimuli, respectively (Camille et al., 2011; Luk and Wallis, 2013; Rudebeck et al., 2008). A causal investigation of their interaction in stimulus-based and action-based learning in rodents is lacking in the broader literature. Further, we recently suggested that a thorough consideration of both accuracy and speed of response during learning may be particularly informative when identifying neural signatures of flexible learning (Aguirre et al., 2020; Harris et al., 2020).

Lesion or inhibition of OFC impairs multiple forms of reward learning and in rats this has most extensively been observed in the reversal phase (Chudasama and Robbins, 2003; Izquierdo et al., 2013), particularly in spatial- or action-based reversal learning (Boulougouris et al., 2007; Dalton et al., 2016; Groman et al., 2019; Riceberg and Shapiro, 2012). Classic recording studies in rat OFC during (olfactory cue) reversal learning (Schoenbaum et al., 1999) show that neural activity represents an integration of the motivational significance of a reward and its specific predictive cue, critical for reversal but not discrimination learning (Marquardt et al., 2017; Schoenbaum et al., 1999). Conversely, lesion or inhibition of ACC results in decrements in (visual) discrimination performance (Chudasama et al., 2003), but have no effect on fully-predictive, deterministic reversal learning (Bussey et al., 1997; Schweimer and Hauber, 2006). Yet ACC neurons in rats integrate motor-related information with expectations of future outcomes (Cowen et al., 2012; Hyman et al., 2017), encode choice value (Mashhoori et al., 2018) and maintain trial-by-trial performance (Akam et al., 2021), all processes that would be vital not just to action-based learning but also to stimulus-based reversal learning. This growing body of evidence supports a general role for ACC in maintaining a model about the reward environment and making predictions about future choices associated with either actions or stimuli.

More importantly, how these two regions interact is unclear. Several groups have recorded oscillatory activity simultaneously in both areas and uncovered overlapping functions, in contrast to the interference studies (Rudebeck et al., 2008; Rudebeck et al., 2006). For example, single neurons in macaque OFC and ACC encode the action that guides choice, with stronger firing patterns in OFC during a stimulus-based task and stronger ACC firing during an action-based task (Luk and Wallis, 2013). More recently, there is reported evidence of ACC-OFC task-related coherence in theta to low beta (4-20 Hz) oscillations as rats respond to or choose a higher-valued reward, suggesting a functional connectivity that encodes relative reward value (Amarante and Laubach, 2021; Fatahi et al., 2020; Fatahi et al., 2018). It is therefore conceivable that these two regions work in a hierarchy for reward-based learning and choice (Hunt and Hayden, 2017; Hunt et al., 2018) potentially with theta oscillations in OFC as the spectral signature that coordinates reward signals (Knudsen and Wallis, 2020; Marquardt et al., 2017) with ACC (Fatahi et al., 2018). However, causal network-level interactions and the spectral signatures in flexible reward learning remain unclear.

Given that ACC is densely innervated by regions involved not only in evaluating options, but also preparing and executing actions, we hypothesized that ACC theta modulates overall

learning in both S-O and A-O tasks. To test this, we investigated the learning of stimulus-outcome (S-O) and action-outcome (A-O) associations using a combination of simultaneous *in-vivo* electrophysiology in dorsal ACC (dorsal area 32/24, van Heukelum et al. (2020)) and ventral OFC, partnered with inhibitory Designer Receptors Exclusively Activated by Designer Drugs (DREADDs) in ACC. We opted to record from posterior VO as it is richly interconnected with ACC (Barreiros et al., 2021; Hoover and Vertes, 2011). We discovered that ACC and OFC theta signaled correct choices in S-O and A-O learning, but had differential roles in discrimination and reversal phases: ACC and OFC theta signaled discrimination accuracy whereas OFC theta alone signaled reversal accuracy. ACC inhibition hastened deliberation speed of incorrect choices in S-O and correct choices in A-O learning, but theta power in either region was not correlated with choice speed, suggesting dissociable roles of these oscillations in modulating accuracy and speed. The present results also point to a specific role for OFC theta in signaling a reversal of either stimuli or actions.

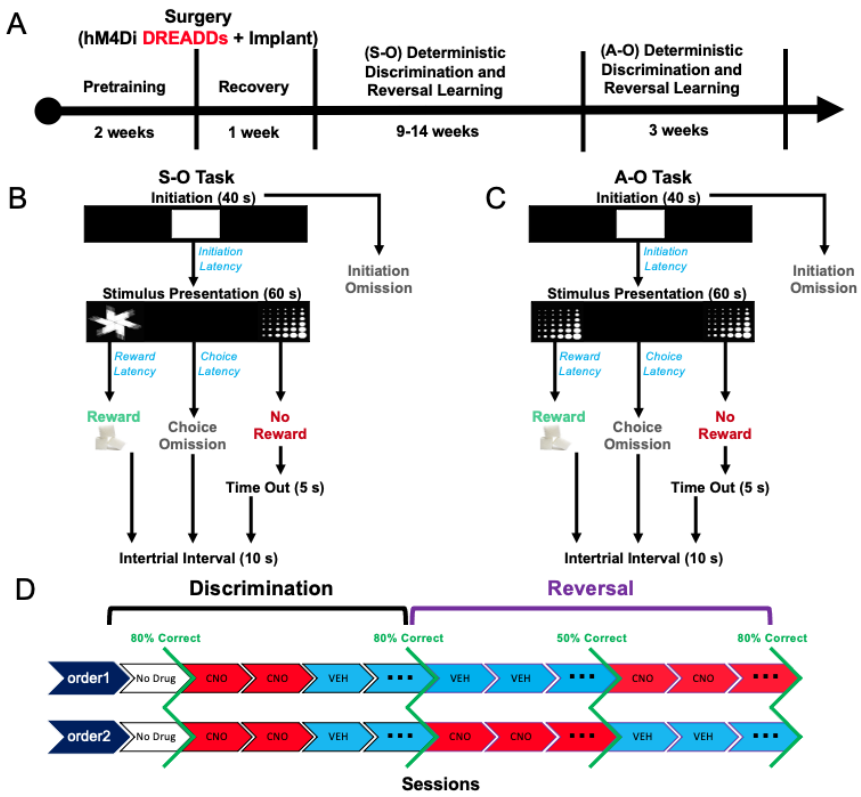


Figure 1. Experimental timeline, task structure, and drug administration. **(A)** After approximately 2 weeks of pretraining, rats underwent bilateral viral infusion of inhibitory hM4Di DREADDs in ACC along with chronic implantation of electrode arrays targeting both ACC and OFC unilaterally in each animal. **(B)** Following a 1-week recovery period, rats began the Stimulus-Outcome (S-O) discrimination/reversal learning task, with rewarded-stimulus assignment (e.g., "fan" vs. "marbles") counterbalanced across rats. **(C)** After completing the S-O task, the rats advanced to an Action-Outcome (A-O) version of the task where successful reward procurement was possible by ignoring stimuli (e.g., "marbles" vs. "marbles") and instead by making a left or right response. A-O stimulus identity was counterbalanced across rats. **(D)** Drug administration (clozapine-N-oxide, CNO or vehicle, VEH) was administered once criterion was reached for initial discrimination to probe the effect of ACC inhibition on mastery-level performance. Subsequently, in the reversal phase, the order of drug administration was counterbalanced across animals.

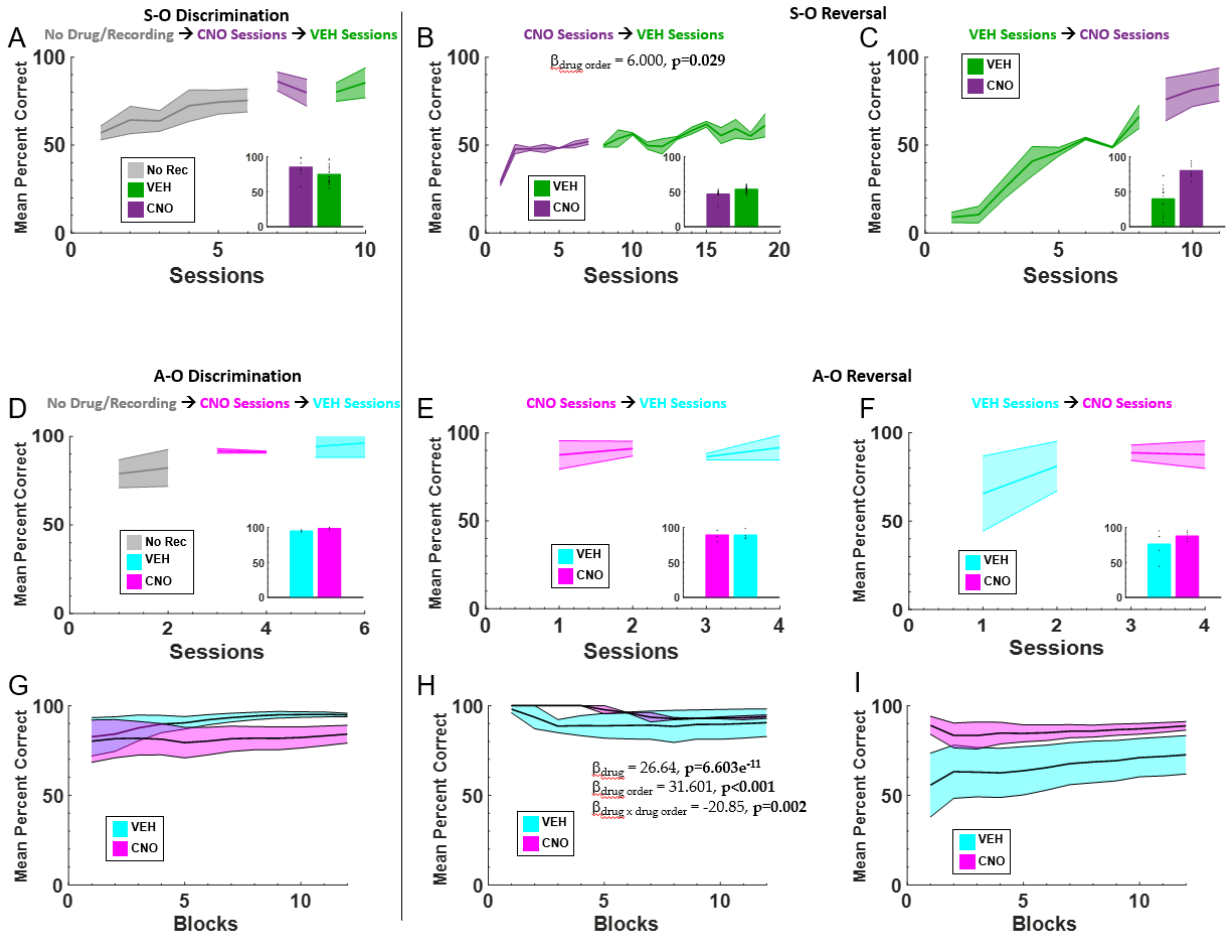


Figure 2. ACC inhibition during early reversal disrupts the learning of S-O and A-O associations. Rats learned S-O associations much more slowly than A-O associations, the latter often within a single session. ACC inhibition attenuates learning in the early reversal phase of both S-O and A-O. (**A**) Learning during S-O initial discrimination for no recording days (No Rec, grey), and sessions preceded by clozapine-N-oxide (CNO, purple) or vehicle (VEH, green) injections. (**B**) Learning during the reversal phase for which animals received CNO prior to reaching 50%, followed by VEH. (**C**) Learning during the reversal phase for which animals received VEH prior to reaching 50% correct, followed by CNO. (**D**) Learning during A-O initial discrimination for no recording days (No Rec, grey), and sessions preceded by clozapine-N-oxide (CNO, magenta) or vehicle (VEH, cyan) injections. (**E**) Learning during the reversal phase for which animals received CNO prior to reaching 50% correct, followed by VEH. (**F**) Learning during the reversal phase for which animals received VEH prior to reaching 50% correct, followed by CNO. ACC and OFC LFPs were recorded on all sessions in which they received injections. Insets: mean percent correct, for each condition. (**G**) Trial-by-trial learning during A-O initial discrimination. (**H**) Trial-by-trial learning during the reversal phase for animals receiving CNO prior to reaching 50% correct, followed by VEH. (**I**) Trial-by-trial learning during the reversal phase for animals receiving VEH prior to reaching 50% correct, followed by CNO. Trial-by-trial plots are sliding averages across twelve blocks with a 3-trial overlap between each block. Means ± standard error of the mean (S.E.M.) shades. Representative trial-by-trial data are shown in **Fig S1**.

RESULTS

ACC supports early S-O and A-O reversal learning

Rats were prepared with inhibitory (hM4Di) DREADDs on putative projection neurons in ACC, implanted with custom-constructed fixed electrode arrays unilaterally in ACC and OFC, and tested on S-O and A-O tasks, using a within-subject design for drug administration (**Figure 1**). Rats were administered clozapine-N-oxide (CNO) once they reached criterion-level performance (80%) on the initial discrimination learning phase followed by CNO, to assess the impact of ACC inhibition on mastery-level performance (Bussey et al., 1997; Chudasama et al., 2003). Subsequently in the reversal phase, they were administered CNO or vehicle (VEH) either early in learning or on reaching criterion-level performance. Given our within-subject, counterbalanced design, we assessed learning as the number of sessions to pass to the next drug stage, which we defined as reaching 50% correct on two consecutive days in reversal, or *sessions to next drug treatment*. We performed a mixed GLM analysis with drug order (CNO-first, VEH-first) as a between-subject factor, drug as within-subject factor, and individual rat as random factor. In S-O reversal learning we found a significant effect of drug order (GLM: $\beta_{drug\ order} = 6.000$, $t(4) = 3.382$ $p=0.029$) and a marginal effect of drug (GLM: $\beta_{drug} = -5.5$, $t(4) = -2.386$, $p=0.075$): Early inhibition of ACC in S-O reversal learning attenuated performance such that it hovered around chance (**Figure 2B**). This learning stands in contrast to performance when VEH was administered

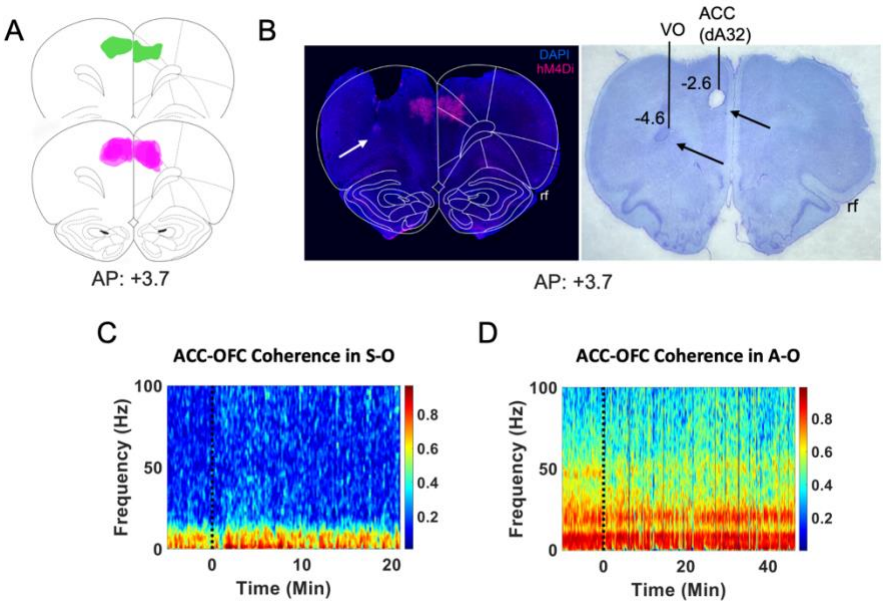


Figure 3. Inhibitory (hM4Di) DREADDs in ACC, electrode placement, and ACC-OFC coherence. (**A**) Reconstructions of placement of inhibitory hM4Di DREADDs (magenta) and eGFP null virus (green) at Anterior-Posterior (AP) level +3.7 relative to Bregma. (**B**) *Left*, Representative placement of inhibitory hM4Di DREADDs at Anterior-Posterior (AP) level +3.7 relative to Bregma. Note, tissue (secondary motor cortex, M2) was removed at time of brain extraction**.** *Right*, Nissl-stained section showing electrolytic lesions and placement of electrode arrays targeting both ACC (dorsal area 32, dA32) and OFC (ventral orbital, VO) unilaterally in a representative animal. **(C)** Representative spectrogram showing ACC-OFC coherence during S-O learning following vehicle injection over an entire session. **(D)** Representative spectrogram showing ACC-OFC coherence during A-O learning following vehicle injection over an entire session, in the same animal as in C. Arrows pointing at tip of electrode tracks. *rf = rhinal fissure.*

first during reversal (**Figure 2C**), where rats instead show the expected tendency to follow the initial (previous) stimulus-reward assignment on early reversal trials during VEH, but still reach higher accuracy levels later in learning when administered CNO. Attenuated early stimulus-

based reversal learning following ACC inhibition is consistent with previous work from our lab (Stolyarova et al., 2019).

Since by comparison with S-O we observed much quicker learning for A-O **(Figure 2D-F),** we analyzed accuracy via blocks of trials (**Figure 2G-I**). Also for A-O, we observed an attenuation of accurate performance with ACC inhibition early in learning, compared to later in learning: We found a significant effect of drug order (GLM: $\beta_{drug\ order}$ = 31.601 t(176) = 3.451 p<0.001), a significant effect of drug (GLM: β_{drug} = 26.64, t(176) = 6.956, p=6.603e^{-11}), and drug order by drug interaction (GLM: $\beta_{drug\ x\ drug\ order}$ = -20.85, t(176) = -3.098, p=0.002). Taken together, the pattern of results points to a time-sensitive role for ACC in support of early reversal learning of either stimuli or actions.

ACC and OFC theta signal accuracy during initial learning

Local field potential (LFP) data in ACC and OFC were acquired at 30 kHz and downsampled to 1,000 Hz. LFPs were bandpass filtered for theta (5-10 Hz) using the *filtfilt()* function in MATLAB. ACC DREADDs placement, electrode placement, and ACC-OFC coherence are shown in **Figure 3**, and average power spectral density in **Figure S7.** Four behavioral task epochs (e.g., trial initiation, correct choice, incorrect choice, and reward port entry) were configured as triggered event outputs from each operant chamber to the data acquisition system that were simultaneously collected with the electrophysiological recordings. Specific events were averaged over trials and over sessions, and subsequently combined across rats. 'Off task' neural data were also obtained following injections while animals were not on a learning task, but simply placed in the operant chamber, which showed no increase in theta power (and no difference between CNO and VEH conditions) for both S-O and A-O. To that end, we performed a random sampling of -2/+2 sec epochs of those 60 min sessions for S-O (**Figures S2E and S2F)** and for A-O (**Figures S3E and S3F)**.

We were primarily interested in measures that we previously operationally defined as measures of accuracy (Harris et al. (2020): choice of correct stimulus, choice of incorrect stimulus, and reward collection).

S-O discrimination learning. For measures where we found significant discrimination and reversal phase interactions (reported in ***Global Measures of S-O learning*** below), we were justified to fit mixed-effects generalized linear models (GLMs) for learning in the discrimination phase separately from the reversal phase. ACC theta, OFC theta, and drug (CNO vs. VEH) were entered into a GLM model as predictors of *choice of the correct stimulus* within the time range of -2 and +2 sec of choice. We found ACC theta exceeded model prediction for this measure during discrimination (GLM: $\beta_{ACCtheta}$ =-35.606, t(19)=-2.346, p=0.03). Post-hoc GLMs revealed the -2 sec pre-choice epoch (β_{pre} = 24.454, t(19)=2.857, p=0.01) and the +2 sec post-choice epoch (β_{post} =20.436, t(19)=-2.707, p=0.01) were both significant. However, the effect of drug was not statistically significant, indicating that theta oscillations in ACC, though decreased following CNO administration **(Figure S2A)**, still predicted choice of the correct stimulus in the discrimination phase (**Figure 4A**). This relationship was not driven by movement speed, as latencies to choose the correct stimulus were not correlated with theta power (**Figure 5**). These effects in the post-

choice epoch (a burst of theta power 1-sec post correct stimulus choice) occur immediately after stimulus choice, but before reward collection, i.e., during reward expectancy.

For *choice of the incorrect stimulus,* we also found a significant effect of ACC theta (GLM: $\beta_{ACCtheta}$ =45.182, t(19)=-2.575, p=0.019). Specifically, theta power in the +2 sec post-choice epoch (β_{post}=24.806, t(19)=2.561, p=0.02) was significant, but the -2 sec pre-choice was not (β_{pre}= 25.863, t(19)=2.057, p=0.05). Further, there was an interaction of ACC theta x OFC theta (GLM: $\beta_{ACCtheta\ x\ OFCtheta}$ =-711.68, t(19)=-2.423, p=0.026) during incorrect choice (**Figure 4C**). Both ACC theta power (t(612)=4.694, p=3.316e^{-06}) and OFC theta power (t(592)=3.354, p=8.479e^{-04}) were significantly greater during *choice of correct* vs. *choice of incorrect stimulus* (**Figure S4**). No effect of ACC inhibition on *choice of the incorrect stimulus* was observed. For *rewards collected*, despite the omnibus across-phase interactions and OFC x ACC theta interactions reported below in ***Global Measures of S-O learning*** analyses, we found no effect of drug, ACC theta, or OFC theta on this measure in the discrimination phase.

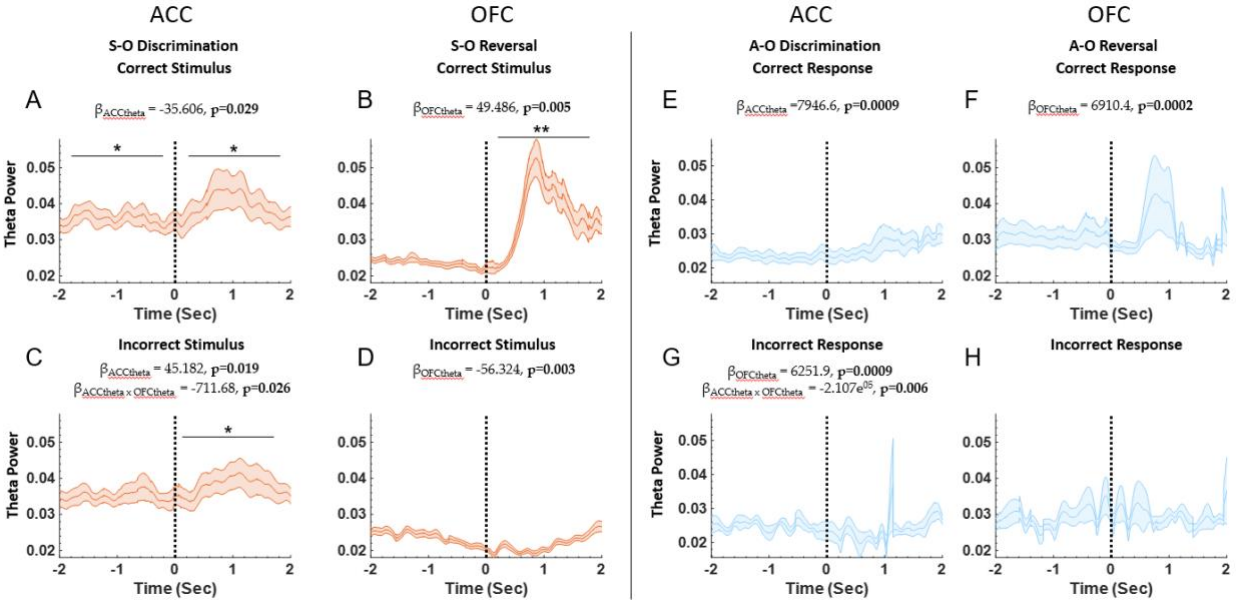


Figure 4. ACC and OFC theta modulate S-O and A-O accuracy in phase-specific ways. (**A,C**) In S-O Discrimination, ACC theta power was a significant predictor of choice of the correct stimulus and choice of the incorrect stimulus. There was also a significant interaction with OFC theta for choice of the incorrect stimulus. (**B,D**) OFC, not ACC, theta power was a significant predictor of choice of the correct stimulus and choice of the incorrect stimulus in S-O reversal learning. There was a significant interaction of ACC theta x drug for the incorrect choice only. (**E,G**) In A-O Discrimination, theta power in both ACC and OFC significantly predicted choice of the correct response. (**F,H**) OFC, not ACC, theta power was a significant predictor of choice of the correct response in reversal learning. There was also a significant interaction of OFC theta x drug. Reported statistics for A-O are for the Sliding Correct measure following trial blocks analysis. Plotted are theta power means ± standard error of the mean (S.E.M.) shading, with clozapine-N-oxide (CNO) and vehicle (VEH) conditions collapsed. For separation between CNO and VEH conditions, see **Fig S2** and **Fig S3**. Post-hoc GLMs for effects at pre- and post- time=0 are noted with significance bars *p<0.05, **p<0.01.

A-O discrimination learning. Given the much quicker learning in A-O compared to S-O, we analyzed behavior by trial block, a finer-grained analysis to session-level performance, for A-O in relation to theta oscillations in ACC and OFC. All animals reached criterion and completed at minimum 60 trials per session (see **Materials and Methods**), we therefore binned the percent

correct of each animal's learning into 12 blocks containing 5 trials each, and calculated a sliding average across all twelve blocks with a 3-trial overlap between each block, for a *Sliding Correct* measure. Theta-filtered LFPs for each trial were also binned and calculated in the same manner and subsequently analyzed using mixed effects GLM, as for S-O. Since the omnibus analyses (***Global Measures of A-O accuracy)*** revealed significant phase interactions, this prompted separate analyses of discrimination and reversal phases, as in S-O above. Identical as above for S-O, ACC theta power, OFC theta power, and drug (CNO vs. VEH) were entered into the GLM model as predictors. During the A-O discrimination phase, factors that exceeded model prediction included ACC theta (GLM: $\beta_{ACCtheta}$=7946.6, t(136)=2.6609, p=0.009), OFC theta (GLM: $\beta_{OFCtheta}$=6251.9, t(136)=2.6627, p=0.009), along with a significant interaction of ACC x OFC theta (GLM: $\beta_{ACCtheta\ x\ OFCtheta}$=-2.107e^{05}, t(136)=-2.7802, p=0.006) for this *Sliding Correct* measure. This relationship was not driven by movement speed, as latencies to make correct or incorrect responses were not correlated with theta power (**Figure 6**). Though *Sliding Correct* was a combined measure obtained for trial block analyses, for direct comparison with S-O learning we show theta power in relation to the specific triggered events, *choice of correct response* and *choice of incorrect response* (**Figures 4E** and **4G**). Unlike S-O, neither ACC theta power (t(336)=0.0260, p=0.979) nor OFC theta power (t(335)=-0.2181, p=0.828) were significantly greater for *choice of correct* vs. *choice of incorrect response* in A-O discrimination learning (**Figure S4**).

Collectively, analysis of the discrimination phase demonstrated that theta power in ACC and OFC signal accuracy in the immediate post-choice epoch during both S-O and A-O initial discrimination learning. However, correct vs. incorrect choices are reflected by significant theta power differences only in S-O, not in A-O task conditions.

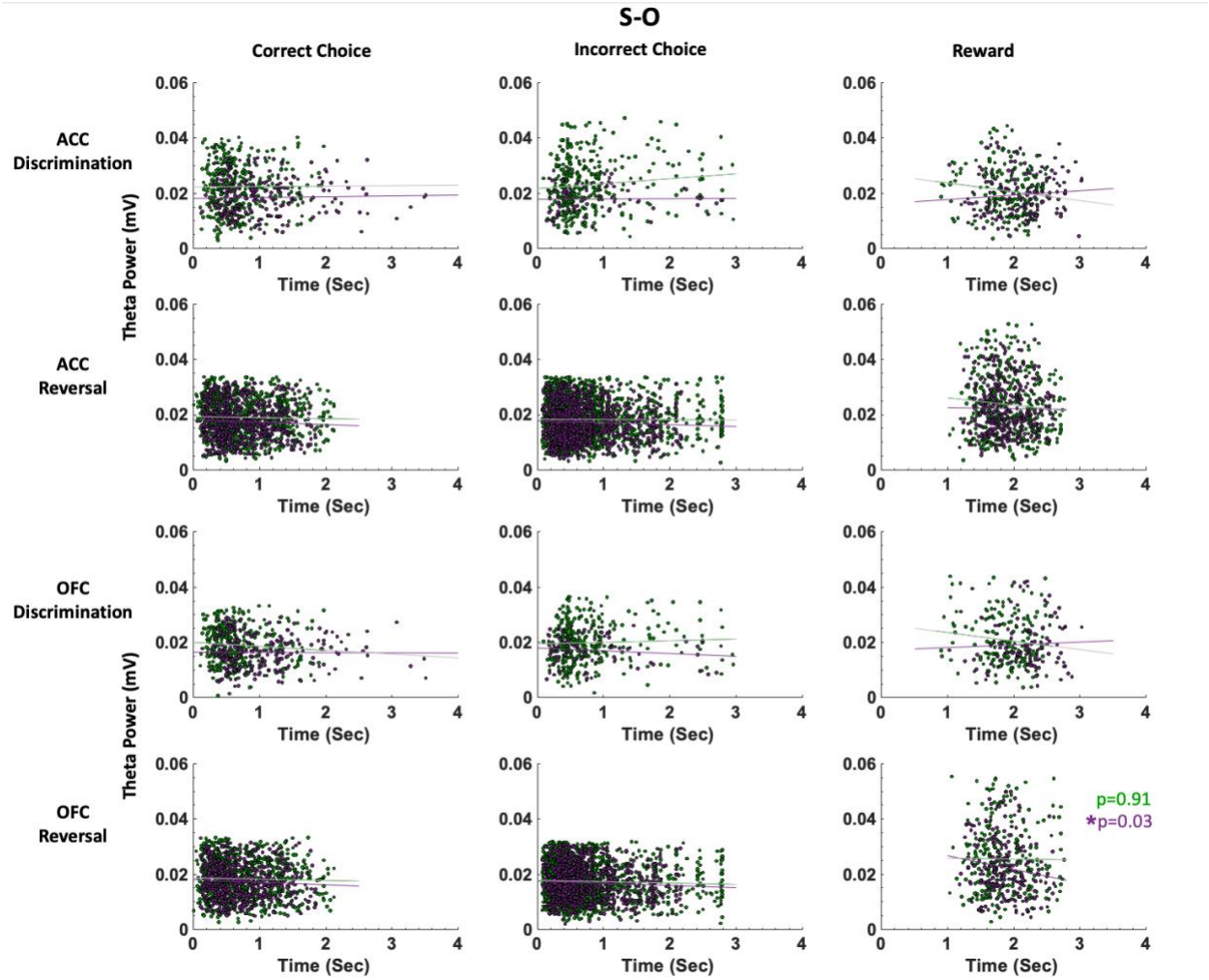


Figure 5. Theta power in ACC and OFC is not correlated with deliberation speed in S-O learning. Theta power was filtered and extracted from ACC and OFC for each event-triggered epoch (Correct choice, Incorrect choice, and Reward port entry). Movement speed was calculated by subtracting the time at the target event (t) from the time at the previous event (t-1) in seconds (Sec). OFC theta power decreased as reward collection latencies increased after CNO injection (purple) during reversal learning. This negative correlation was not observed following VEH injection (green). There were no other significant correlations between theta power and choice of correct or incorrect stimulus. *p<0.05.

OFC theta alone signals accuracy during reversal learning

S-O reversal learning. For the reversal phase, ACC theta, OFC theta, and drug (CNO vs. VEH) were entered into a GLM model as predictors of performance. We found consistent evidence of an OFC theta signal of accuracy: For *choice of the correct stimulus*, we found a significant effect of OFC theta alone (GLM: $\beta_{OFCtheta}$ =49.486, t(67)=2.886, p=0.005). Specifically, theta power in the +2 sec post-choice epoch (β_{post}=48.488, t(67)=2.72, p=0.008) was significant, and not the -2 sec pre-choice epoch (β_{pre}= 0.749, t(67)=0.037, p=0.97), suggesting OFC theta signals reward expectancy even on the relatively fewer correct trials in reversal compared to incorrect trials (**Figure 5**). The effect of drug was not statistically significant, indicating that theta oscillations in OFC, though decreased by CNO (**Figure S2C**), still modulated choice of the correct stimulus in

the reversal phase, and this was driven by the post-choice epoch (**Figure 4B**). Importantly, movement speed during choice of the correct stimulus during reversal was not correlated with OFC theta power (**Figure 5**). For *choice of the incorrect stimulus,* we once again found a significant effect of OFC theta (GLM: $\beta_{OFCtheta}$ =-56.324, t(67)=-3.130, p=0.003). However, post-hoc analyses of pre and post choice epochs revealed that theta power in the post-choice and pre-choice epochs was not a significant predictor. Further, we found a significant interaction of drug x ACC theta (GLM: $\beta_{drug\ x\ ACCtheta}$ =-69.883, t(67)=-1.998, p=0.049) for this trial epoch alone (**Figure 4D**). Post-hoc analyses with ACC theta power as a predictor in the model, however, did not reveal a significant differential effect of CNO (p=0.83) vs. VEH (p=0.72). And finally, for *rewards collected*, despite the omnibus across-phase interactions and OFC and ACC theta interactions reported in the ***Global Measures of S-O learning*** section, we found no effect of drug, ACC theta, or OFC theta power on this measure in the reversal phase. Taken together, analysis of the S-O reversal phase demonstrated that OFC, not ACC, theta power signals accuracy, and that choice of the incorrect stimulus is modulated by drug (i.e., ACC inhibition). OFC theta power was significantly greater during *choice of correct* vs. *choice of incorrect stimulus* (t(467)=5.547 p=4.874e^{-08}) during S-O reversal learning, **Figure S4**.

A-O reversal learning. For A-O *Sliding Correct* in the reversal phase, as in S-O learning, only OFC theta significantly signaled this measure (GLM: $\beta_{OFCtheta}$=6910.4, t(171)=3.8077, p=0.0002), **Figure 4F** and **4H**). There was also a significant drug x OFC theta interaction (GLM: $\beta_{drug\ x}\ \beta_{OFCtheta}$ =-7729.3, t(171)=-3.986, p=9.932e^{-05}).

This finer-grained trial block analyses revealed a near-identical pattern in A-O to the S-O results: Both ACC and OFC theta signal accuracy in the discrimination phase, whereas OFC theta alone signals accuracy during reversal. ACC and OFC theta power for trial epochs and interaction with drug are shown in **Figure S3.** However, an appreciable difference from S-O was that for A-O reversal learning OFC theta power was not significantly greater for *choice of correct* vs. *choice of incorrect response* (t(478)=1.632, p=0.103). Here again as in the discrimination phase, we find that correct vs. incorrect choice is reflected by theta power differences only in S-O, not in A-O task conditions (**Figure S4**). Notably, the role of OFC theta in signaling accuracy during these trials was not driven by movement speed, as latencies to make a correct or incorrect choice were not correlated with theta power in either the S-O (**Figure 5**) or A-O (**Figure 6**) reversal phase.

Our results confirm and extend a recent report that ACC-OFC theta coherence encodes reward value (Amarante and Laubach, 2021), since the ACC-OFC interaction we observed occurred during initial learning of both S-O and A-O tasks, but not during reversal learning. A noteworthy difference from Amarante & Laubach (2021) is that this group studied coherence during reward consumption (i.e., licks). We note that the increase in theta power we report here occurred immediately following a correct choice but preceded most reward collection times (1.5-2.5 sec) and consumption of the sucrose pellet. Reward consumption (and a chewing artefact) would appear uniformly across all LFP channels and both regions, whereas the ACC and OFC signals were not uniform here. Interestingly, though ACC inhibition decreased theta power during discrimination learning, and with CNO delivered at a dose that produced behavioral impairments on two different operant tasks in our lab previously (Hart et al., 2020; Stolyarova et al., 2019), this inhibition was not enough to disrupt the theta signal of accuracy.

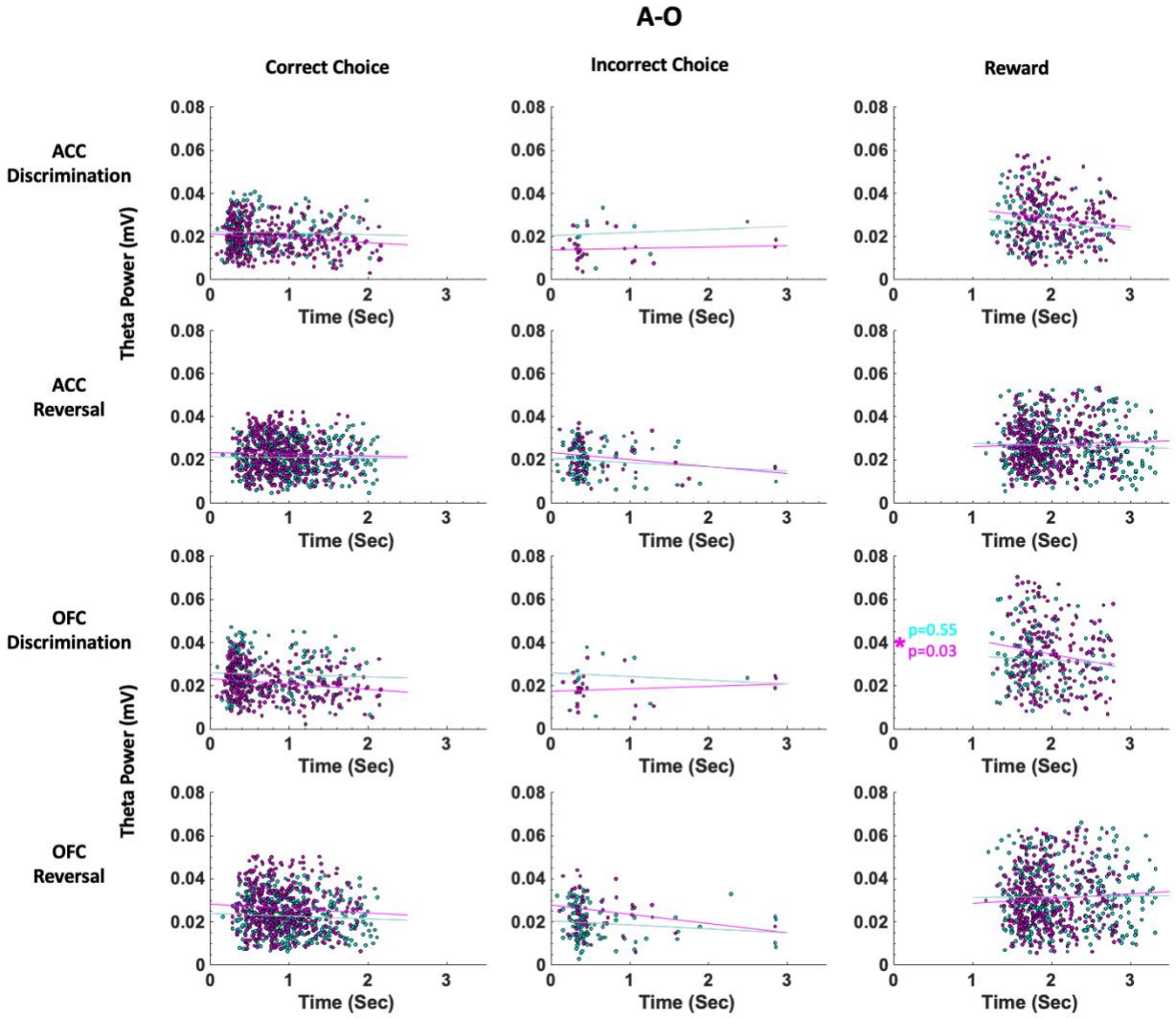


Figure 6. Theta power in ACC and OFC is not correlated with deliberation speed in A-O learning. Theta power was filtered and extracted from ACC and OFC for each event-triggered epoch (Correct choice, Incorrect choice, and Reward port entry). Movement speed was calculated by subtracting the time at the target event (t) from the time at the previous event (t-1) in seconds (Sec). OFC theta power decreased as reward collection latencies increased after CNO injection (magenta) during discrimination learning. This negative correlation was not observed following VEH injection (cyan). This was an identical pattern to that observed in S-O reversal learning. There were no other significant correlations between theta power and choice of correct or incorrect response. *p<0.05.

ACC inhibition hastens deliberation speed for incorrect choices in S-O reversal learning: A mechanism for impairment

Because we were also interested in theta oscillations as they relate to speed of response, we next analyzed median latencies, i.e., to initiate a trial, to choose the correct stimulus, to choose the incorrect stimulus, and to collect reward (Harris et al., 2020). We further probed measures for which the omnibus analyses (***Global Measures of S-O speed***) revealed significant phase and drug interactions: *incorrect choice latencies* and *reward collection latencies*.

For the discrimination phase, there was no significant effect of drug, ACC theta, or OFC theta on these measures for *incorrect choice latencies*. There was a significant interaction of drug x OFC theta (GLM: $\beta_{drug \times OFCTheta}$ =-831.12, t(19)=-3.131, p=0.006) on *reward collection latencies* (**Figure 7**) but post-hoc analyses revealed no differential effect of CNO (p=0.63) vs. VEH (p=0.75), with OFC theta power as predictor in the model.

For the reversal phase, we similarly analyzed measures for which the omnibus analyses revealed significant phase and drug interactions: *incorrect choice latencies* and *reward collection latencies*. For *incorrect choice latencies*, we found a significant effect of drug (GLM: β_{drug} =-4.062, t(67)=-2.289, p=0.025; VEH slower than CNO, respective means 1.88s vs 1.72s), and significant effect of OFC theta (GLM: $\beta_{OFCtheta}$ =-89.086, t(67)=-2.377, p=0.020), indicating greater theta power associated with shorter latencies on incorrect choices **(Figure 7)**. For *reward collection latencies* in the reversal phase, we also found a significant effect of drug (GLM: β_{drug} =-4.217, t(67)=-2.081, p=0.041; VEH slower than CNO, respective means 0.63s vs 0.52s). OFC theta power decreased as reward collection latencies increased following CNO. A significant negative correlation between OFC theta power and reward collection speed (r=-0.14934, p=0.031) was nonsignificant in the VEH condition (r=-0.0071, p=0.911; **Figure 5**).

Collectively, the analyses of latencies in S-O learning reveal a quickening of deliberation speed for incorrect trials following CNO compared to VEH during the reversal phase, a likely contributor to a reversal learning impairment. Further, ACC inhibition changes the relationship between OFC theta power and reward collection speed, also during the reversal phase. Critically, the reward collection trial epoch is most associated with sensitivity to, or changes in, motivation (Harris et al., 2020). We show CNO does not produce global locomotor effects which would be reflected across all trial epochs, as may be expected from a previous report (MacLaren et al., 2016).

ACC inhibition quickens deliberation speed for correct choices in A-O learning: More interaction with ACC and OFC theta

In the discrimination phase, there was a significant effect of drug on *correct choice latencies* where CNO also quickened responses (GLM: β_{drug} =27.736, t(5)=2.711, p=0.042; VEH slower than CNO, respective means 0.98s vs 0.79s), **Figure 7**. There was also a significant drug x OFC theta interaction (GLM: $\beta_{drug \times OFCtheta}$ =-1149.4, t(5)=-3.341, p=0.021), yet post-hoc tests with OFC theta power as a predictor in the model revealed no differential effect of CNO (p=0.46) vs. VEH (p=.21) during this trial epoch. Similarly, there was a drug x OFC theta x ACC theta interaction (GLM: $\beta_{drug \times OFCtheta \times ACCtheta}$ =33276, t(5)=3.265, p=0.022) in this trial epoch, but post-hoc tests revealed that OFC theta power in either CNO (p=0.68) or VEH conditions (p=0.25) was not predictive of correct choice latencies, nor with ACC theta power as predictor in the model (CNO, p=0.89 vs. VEH, p=0.97). There was a marginal interaction of drug x ACC theta (GLM: $\beta_{drug \times ACCtheta}$ =-750.88, t(5)=-2.447, p=0.058) which we did not follow-up with post-hoc tests**.** Though we found a significant phase x drug interaction in the omnibus test for *initiation latencies* (reported in ***Global Measures of A-O speed***, below), further analysis revealed that OFC theta was not a significant predictor of this variable.

We found significant model prediction of *correct choice latencies* by drug, ACC theta, and OFC theta (and significant interactions) in the reversal phase. Once again, there was a significant

effect of drug (GLM: β_{drug} =9.584, t(7)=4.697, p=0.002; VEH slower than CNO, respective means 0.90s vs 0.74s), ACC theta (GLM: $\beta_{ACCtheta}$ =387.71, t(7)=4.886, p=0.002), and OFC theta (GLM: $\beta_{OFCtheta}$ =284.07, t(7)=3.378, p=0.012) for this trial epoch. We found a significant interaction of drug x OFC theta (GLM: $\beta_{drug\ x\ OFCtheta}$ =-333.56, t(7)=-3.947, p=0.006); post-hoc tests with OFC theta power as a predictor in the model revealed a significant effect of both CNO (p<0.01) and VEH (p<0.01) during this trial epoch. We also found a significant drug x ACC theta interaction (GLM: $\beta_{drug\ x\ ACCtheta}$ =-427.87, t(7)=-5.255, p=0.001); post-hoc tests with ACC theta power as a predictor in the model revealed a significant effect of CNO (p<0.01) but not VEH (p=0.25) during this trial epoch. There was also a significant interaction of ACC x OFC theta (GLM: $\beta_{ACCtheta\ x\ OFCtheta}$ =-13587, t(7)=-4.172, p=0.004), and a 3-way interaction of drug x ACC theta x OFC theta (GLM: $\beta_{drug\ x\ ACCtheta\ x\ OFCtheta}$ =14549, t(7)=4.452, p=0.003): post-hoc tests revealed that OFC theta power in either CNO (p=0.02) or VEH conditions (p=0.01) was predictive of *correct choice latencies*, but not with ACC theta power as predictor in the model (CNO, p=0.23 vs. VEH, p=0.12).

Collectively, these analyses suggest the effects of ACC inhibition on *correct choice latencies* during A-O reversal learning, unlike S-O, were significantly modulated by OFC and ACC theta

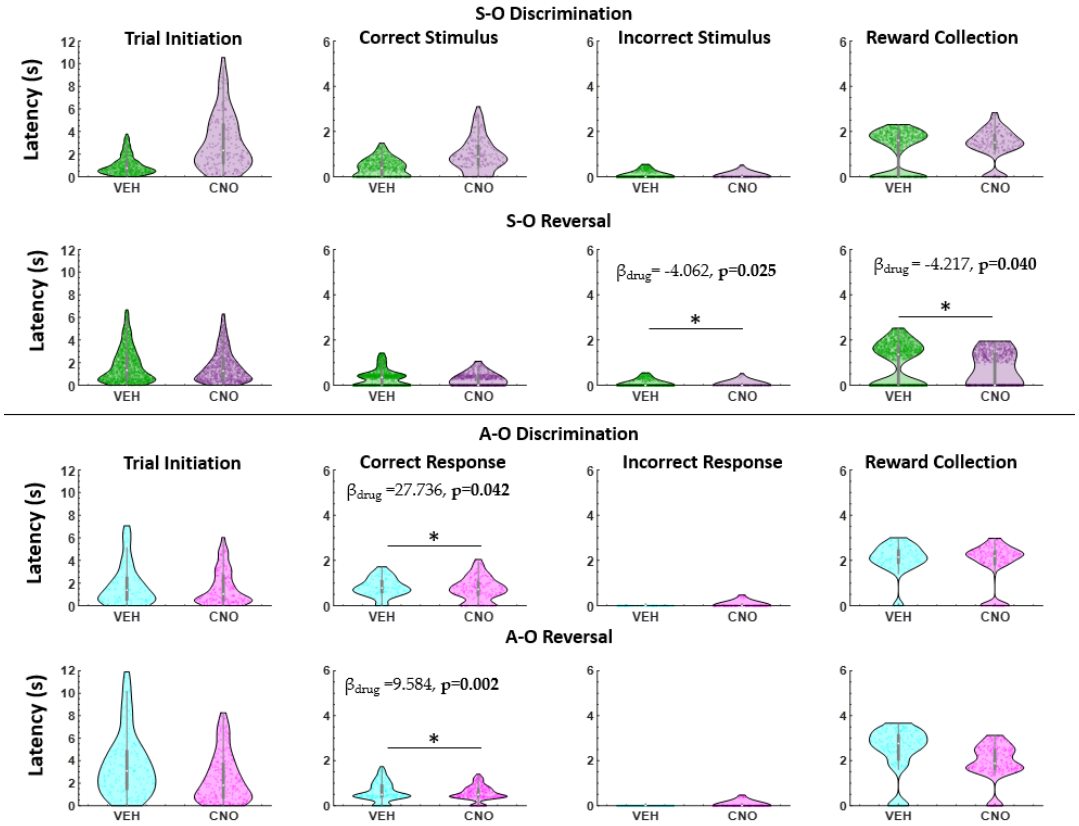


Figure 7. ACC inhibition decreased deliberation speed during discrete S-O and A-O trial epochs. Violin plots depicting individual session median latencies in seconds (s) for each of the trial epochs in discrimination (top) and reversal learning (bottom). Individual dots represent trials. Clozapine-N-oxide (CNO) administration significantly reduced the latency to choose the incorrect stimulus and collect reward in S-O learning. CNO administration significantly decreased the latency to choose the correct stimulus in A-O learning. Despite its prominent role in accuracy in the discrimination phase of S-O, ACC theta did not predict speed of responses for any trial epoch in either the discrimination or reversal phase. In contrast, both ACC and OFC theta significantly predicted speed of correct choice during A-O reversal learning (all significant Beta values are not shown, but are reported in text). *p<0.05, **p<0.01 CNO vs. VEH.

power. The overall effect of ACC inhibition was a hastened deliberation on correct trials. A summary of all mixed-effects GLM post-hoc analyses can be found in **Table 1**.

DISCUSSION

There have been extensive reports on specializations of OFC and ACC in stimulus- and action- based learning, respectively, in macaques and humans (Camille et al., 2011; Luk and Wallis, 2013; Rudebeck et al., 2008). Here we report that ACC and OFC *both* participate in stimulus- and action-based learning in rat, albeit in different ways. In the present study we were specifically interested in theta oscillations in frontal cortex since there have been several reports of involvement of oscillations in this frequency band in reward value and decision making in rodents (Amarante and Laubach, 2021; Fatahi et al., 2020; Fatahi et al., 2018; Marquardt et al., 2017), and one report showing that theta oscillations in OFC are critical for reversal learning in non-human primates (Knudsen and Wallis, 2020). Specifically, we found significant ACC and OFC theta signaling of reward expectancy (i.e., post-correct choice) during the initial learning of stimuli and actions. Notably, visual stimulus-based learning paradigms (Chen et al., 2021; Harris et al., 2020; Hervig et al., 2020) are less frequently administered in freely-moving rodents compared to paradigms involving action (spatial)-based methods wherein animals nosepoke left/right locations or press left/right levers. Plots of trial-by-trial performance (**Figure S1**) confirm that animals often learned A-O associations within-a-session, whereas animals learn S-O associations across sessions, exhibiting only small gains from one session to the next. Of course, had the experiments involved odor stimuli (Schoenbaum et al., 1999), we would expect much quicker learning. Nevertheless, there are merits to studying visual stimulus-based learning as it is similar to non-human primate studies where learning in that modality is the norm and spatial/action-based responses are infrequently imposed, thus filling a gap in the comparative literature (Rudebeck and Izquierdo, 2021).

To the extent that we can make species comparisons, damage to macaque ACC (area 24) attenuates action-based reversal learning as outlined above, but there is also a small number of reports of deficits in stimulus-based tasks following ACC lesions, though not reversal learning *per se* (Chudasama et al., 2013; Kennerley et al., 2006). Additionally, marmoset perigenual ACC (area 32) and OFC (area 11 and 13) are both required for visual stimulus-reward contingency learning (Jackson et al., 2016), further indicating that there may be some overlap of OFC and ACC support of stimulus-based learning in primate frontal cortex. Here we provide to our knowledge the first evidence of a similar "shared" function for OFC and ACC in rats.

Our observation that ACC-OFC theta power was a significant predictor of discrimination learning accuracy and that OFC theta power was a significant predictor of reversal learning accuracy accords well with prior lesion work (Bussey et al., 1997; Chudasama et al., 2003) indicating that there is more specialization of ACC for "new" learning and OFC for reversal learning. In the present set of experiments, ACC theta power was a significant predictor of accuracy, strongly involved in signaling the correct choice in the initial phase of both S-O and A-O learning. Indeed, it has previously been proposed that ACC is where learning rates are set (Behrens et al., 2007; Bryden et al., 2011; Totah et al., 2009). In support of this idea, neurons in medial frontal cortex (area 24/25, partially encompassing ACC) bias choices with slow decay in

mice (Bari et al., 2019), suggesting this region generates and maintains stable value representations. Similarly, ACC may not only signal reward prediction errors but also the need to attend to trials following errors (Bryden et al., 2011), which is critically important to task performance monitoring (Hayden et al., 2011). In contrast, we found OFC to be more selectively involved in signaling accuracy during the reversal phase in both S-O and A-O learning (both trial epochs of correct and incorrect choice). This, in turn, is consistent with reports of a more selective involvement of OFC theta signaling updates, not establishment, of initial values during learning (Marquardt et al., 2017). This also accords well with lesion and inactivation studies of S-O learning (Chudasama and Robbins, 2003; Izquierdo et al., 2013). Taken together, we report to our knowledge the first evidence that OFC signals accurate choices during either S-O or A-O reversals in rats (Rudebeck and Izquierdo, 2021).

There were more effects of ACC inhibition and a more prominent role of OFC theta on choices latencies in learning. ACC DREADDs inhibition quickened reward collection speed and incorrect choice latencies, the latter providing a mechanism for the disrupted early S-O reversal learning, which is an effect we have reported before (Stolyarova et al., 2019). Theta oscillations in both ACC and OFC appeared to be more thoroughly integrated into response speed during A-O learning: this was especially true for speed of correct responses (**Figure 7**). In contrast to the selective (and complementary) involvement of ACC and OFC theta in S-O learning accuracy, we instead found strong evidence of coordinated accuracy *and* speed in A-O learning by ACC and OFC theta (**Table 1**). Interestingly, unlike the pattern where ACC inhibition decreases theta power in S-O, ACC inhibition produces hypersynchronous, enhanced theta power in OFC during A-O reversal learning (**Figures S3 and S4**). Taken together with the quicker latencies of correct responses following CNO, ACC inhibition may in part facilitate OFC theta signaling of A-O accuracy and speed. Previous findings of impaired reversal learning after chemogenetic manipulations of either ACC or OFC should be revisited for possible effects on deliberation speed (not a general effect on locomotor speed), as this could be the source of poor performance.

It is possible that expertise in S-O learning influenced A-O learning and rendered animals better able to integrate accuracy and speed of response; effectively 'learning to learn' (Faraut et al., 2016). Because we did not counterbalance the order of S-O and A-O here, extensive experience with S-O associations could have introduced more of a strategy shifting component upon commencement of the A-O task, which would have recruited and engaged more medial sectors of rat frontal cortex (Floresco et al., 2008). Indeed, set shifting is thought to involve gamma oscillations, specifically theta-gamma coupling, in frontal cortex (Voloh et al., 2015). The possibility that we engaged a strategy or set-shifting process is unlikely since there was no pronounced effect of ACC inhibition on A-O discrimination, when one would expect a "set shift" effect. Additionally, we theorized based on the available literature that rat ACC would be the driver of both S-O and A-O learning, which provided the rationale for placing DREADDs in ACC and testing its involvement causally. Equally plausible, however, is that OFC has a different mechanistic role compared to ACC, which remains to be tested. Finally, the influence of probabilistic outcomes could change the nature of these interactions, and involvement of theta. In our effort to equate S-O and A-O tasks as much as possible (i.e., imposing the same criterion, same number of reversals), animals committed many fewer errors in the A-O discrimination phase than in any other phase. Thus, an important follow-up line of work would be to provide

443 more equivalent trial numbers across the A-O and S-O tasks, which could be possible if the tasks
444 were more difficult by introducing probabilistic outcomes. These are all improvements and
445 avenues of future studies.
446 In conclusion, we find similar roles for ACC and OFC across S-O and A-O learning, and
447 report their engagement is dependent on phase (i.e., discrimination or reversal). In more general
448 terms, ACC-OFC interactions are critical in accurately monitoring reward contingencies and
449 learning of the task initially, but not during reversal. Further, neural activity in OFC is more
450 predictive of, and involved in, reversal learning accuracy in both S-O and A-O learning (a reversal
451 of either kind) especially when ACC is taken offline, suggesting complementary roles for ACC
452 and OFC in flexible learning.
453
454
455 **MATERIALS AND METHODS**
456

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
Recombinant DNA Reagent	pAAV8-CaMKIIa-EGFP	Addgene	Cat: 50469-AAV8	Lot # v24783
Recombinant DNA	pAAV8-CaMKIIa-hM4D(Gi)-mCherry	Addgene	Cat: 50477-AAV8	Lot # v95691 Lot # v13094
Chemical compound, drug	Clozapine *N*-Oxide	Tocris Bioscience	Cat: 4936	Batch #: 13B/220034
Chemical compound, drug	Clozapine *N*-Oxide	Research Triangle Institute	Ref #: 032301	RTI Log #: 13626-120
Chemical compound, drug	Clozapine *N*-Oxide	Hellobio	Cat: HB1807-100mg	Freebase, Batch: E0699-2-4
Chemical compound, drug	Clozapine *N*-Oxide	NIDA drug supply program	Reference #: 032301	Freebase RTI Log #: 13626-120

SURGERY				
Other	Stereotaxic Instrument	Kopf	Model 942	
Other	Isoflurane machine	Patterson Veterinary	Product #: 78931132	
Other	Fluriso (Isoflurane, USP)	VetOne	Item #: 501017	250 mL/bottle
Other	Microscope	AmScope	SKU: SM-4B-80S	7x-45x Magnification
Other	Stereotrode Wire (Tungsten 99.95#)	California Fine Wire	MO #: M546910	25uM diameter
Other	Anchor Screws	McMaster-Carr	Cat #: 18-8SS	Fillister head 0-80 x 1/8″ Stainless Steel
Other	Microinjector Pump	Stoelting	Model: 780200S	
Other	Microinjection Syringe	Hamilton	Model: 801RN	Lot #: 739025
Other	Microinjection Needle	Hamilton	Model: 26s gauge/2″/3	Lot #: 508251
Other	Guide Cannula	PlasticsOne	Part: 8IC315GXXXXC	11mm length Lot #: J407241
Other	Injection Cannula	PlasticsOne	Part: 8IC315IXXXXC	12mm length Lot #: J407562
Other	Polyethelene Tubing	Instech Laboratories Inc.	Part: BTPE-20	.38 x 1.09mm diameter
Other	Heat Therapy Pump	Stryker Medical	Model #: TP700	

Other	Euthanasia	VetOne	Item #: 501016	100 mL/bottle
Other	Antibiotic (Sulfatrim Pediatric Suspension)	Pharmaceutical Associates, Inc.	Item #: NDC 0121-0854-16	16 oz bottle
Other	NSAID (Carprofen)	Rimadyl	NADA#: 141-199	50 mg/mL in 20 mL bottle
Other	10% Buffered Formalin Acetate	Fisher Chemical	Item #: SF99-20	Lot #: 181457
DATA				
Other	Bussey-Saksida Rodent Touchscreen Chambers	Lafayette Instruments	Model: 80604EH	
Other	Sucrose Dustless Precision Pellets	Bio-Serve	Product #: F0021	45 mg
Other	Electrophysiology Acquisition System	Blackrock Microsystems	Model: CerePlex Direct	
Other	Electrophysiology Commutator	Dragonfly Inc.	Model #: SL-10-C-BR	
HISTOLOGY				
Biological Stain	Cresyl Violet acetate	Sigma Aldrich	SKU: C5042-10G	Lot #: MKCL2973

Primary Antibody	Anti-mCherry Antibody Pre-Adsrobed (Rabbit Polychlonal)	Rockland Immunochemicals , Inc..	Ref: 600-401-P16	Lot #: 46705
Secondary Antibody	Alexa Fluor 594 goat anti-rabbit lgG (H+L)	Invitrogen (Thermo Fischer Scientific)	Ref: A11037	Lot #: 2079421
Anti-fade reagent	ProLong Gold antifade reagent with DAPI	Invitrogen (Thermo Fischer Scientific)	Ref: P36931	Lot #: 2273701
ANALYSIS				
Software, algorithm	MATLAB	MathWorks	RRID:SCR_001622	Version: 2020b
Software, algorithm	Photoshop	Adobe	RRID: SCR_014199	Version 22.4.2
Software, algorithm	Illustrator	Adobe	RRID:SCR_010279	Version 25.2.3
Software, algorithm	Excel	Microsoft	RRID: SCR_016137	
Software, algorithm	ABET II Touch	Lafayette Instruments	RRID: SCR_020249	

Animals

A total of N=8 Long-Evans rats (Charles River Laboratories, Hollister, CA) was used for these experiments. Three animals were not included in the final dataset due to chronic implant failure. Thus, the final numbers include male (*n*=2) and female (*n*=3) animals that were post-natal day 40 (~250g) upon arrival and were pair-housed in a 12-h reverse light/dark cycle room (lights on at 06:00h), maintained in 22-24ºC temperature conditions, with food and water available *ad libitum* prior to behavioral testing. After one week of acclimation to our vivarium, all animals were handled for 10 min in pairs for 5 days. After the handling period, animals were individually-housed to carefully monitor food consumption under restricted access. All procedures (**Figure 1**) were in accordance with the recommendations of the Guide for the Care and Use of Laboratory

468 Animals of the National Institutes of Health. The protocol was approved by the Chancellor's
469 Animal Research Committee at the University of California, Los Angeles.

470

471 *Viral Constructs*
472 An adeno-associated virus AAV8 driving the hM4Di-mCherry sequence under the CaMKIIa
473 promoter was used to express DREADDs on putative projection neurons in ACC (*AAV8-*
474 *CaMKIIa-hM4D(Gi)-mCherry*, packaged by Addgene, Addgene, viral prep #50477-AAV8). A virus
475 lacking the hM4Di DREADD gene and containing the fluorescent tag EGFP (*AAV8-CaMKIIa-*
476 *EGFP*, packaged by Addgene) was infused into ACC in one animal as a null virus control. The
477 DREADDs-transfected animals and null EGFP animal underwent identical surgeries, and all of
478 our behavior was conducted in a counterbalanced within-subject design with each animal serving
479 as its own control (VEH compared to CNO). Collectively, this allowed us to control for non-
480 specific effects of surgical procedures, exposure to AAV8, and non-specific effects of CNO
481 (detailed below). We note that several recent electrophysiology experiments combining neural
482 recordings with DREADDs do not include a virus (eGFP) control (Alexander et al., 2018; Liu et
483 al., 2017; Schmidt et al., 2019), as we do here.
484

485 *Behavioral Apparatus*
486 Behavioral testing was conducted in operant conditioning chambers shielded for
487 electrophysiological recordings and outfitted with an LCD touchscreen opposing a sucrose pellet
488 dispenser (Lafayette Instrument Co., Lafayette, IN). Sucrose pellets (45mg, Dustless Precision
489 Pellets #F0023, Bio-Serv) were used as rewards. The inner chamber walls were modified to
490 accommodate implanted animals by maximizing the space available (24cm x 33.2cm x 39.5cm)
491 and to improve electrophysiological signal quality. All chamber equipment was controlled by
492 customized ABET II TOUCH software.

493

494 *Pretraining*
495 Immediately following the 5 days of handling, rats were placed on food restriction (12
496 g/d) to no less than 85% of their free-feeding body weight throughout the entire experiment and
497 were weighed each day and monitored closely to not fall below this percentage. Animals were
498 pretrained using the protocol as previously published (Stolyarova and Izquierdo, 2017). Briefly,
499 a series of pretraining stages: Habituation, Initiation Touch to Center (ITC), and Immediate
500 Reward (IM), were designed to train rats to nosepoke, initiate a trial, and select a stimulus to
501 obtain a reward.
502 In the habituation stage, five pellets were automatically dispensed and rats were required
503 to consume all pellets within 15 min to advance. In the ITC stage, the center of the touch screen
504 displayed a white square graphic against a black background. A single sucrose pellet was
505 dispensed with simultaneous onset of audio tone and illumination of the reward receptacle if the
506 rat nosepoked the white stimulus, if rats took longer than 40 s to take this action, the white
507 stimulus disappears without reward and trial was considered an omission followed by a 10 s
508 inter-trial interval (ITI). Rats were required to reach a criterion of consuming 60 rewards in 45

min in order to advance to IM training. In this stage, the same white stimulus was presented as in ITC and a nosepoke indicated trial initiation. The disappearance of the white stimulus was immediately followed by the presentation of a target stimulus on the left or right side of the touchscreen (i.e., forced choice), nose poking this target stimulus was paired with a reward. An *initiation omission* was scored if rats took longer than 40 s to initiate the trial, a *choice omission* was scored if rats failed to nosepoke the left or right stimulus after 60 s. A criterion of 60 rewards consumed in 45 min across two consecutive days was required to advance to the discrimination phase of the behavioral task.

Surgical Procedures

After completing the final pretraining stage, rats were anesthetized with isoflurane for bilateral ACC DREADDS (AAV8-CaMKIIα-hM4D(G_i)-mCherry, Addgene, Cambridge, MA, Addgene, viral prep #50477-AAV8) infusion and unilateral chronic implantation of electrode arrays in both ACC and OFC. Craniotomies were created and a 26-gauge guide cannula (PlasticsOne, Roanoke, VA) were lowered into ACC (AP: +3.7, ML: ±0.8, DV: -2.6 from skull surface), after which a 33-gauge internal cannula (PlasticsOne, Roanoke, VA) was inserted. This set of coordinates for ACC was the more anterior site of the two sites that our lab has previously utilized: more posterior ACC has been probed in effort-based tasks (Hart et al., 2020) whereas stimulus-based reversal learning has been probed with this more anterior targeting (Stolyarova et al., 2019), constituting area 32 of cingulate cortex (van Heukelum et al., 2020). Animals were infused with 0.3 μL of virus at a flow rate of 0.1 μL/minute, with cannulae left in place for 5 additional minutes to allow for diffusion.

In the same surgery after DREADDs infusion, rats were implanted with a custom-designed 16-channel electrode array manufactured in-house. Each array was composed of 8 stereotrodes (California Fine Wire Co., Grover Beach, CA) with four targeting the ACC and another four aimed simultaneously in the ipsilateral OFC (AP: +3.7, ML: ±2.0, DV: -4.6, from skull surface), **Figure 3A**. Electrodes were lowered into the same coordinates for ACC as the viral infusions and fixed in place. Stainless steel anchor screws were placed across the skull surface and a ground/reference screw was inserted over the cerebellum. Arrays were secured in place with cyanoacrylate, bone cement (C&B Metabond, Parkell Inc., Edgewood, NY), and dental acrylic (Patterson Dental, St. Paul, MN). Animals were provided with post-operative Carprofen (5 mg/kg, *s.c.*; Zoetis, Parsippany, NJ), topical anti-biotic ointment (Water-Jel Technologies, Carlstadt, NJ) at the incision site, and oral antibiotics (Sulfamethoxazole and trimethoprim oral suspension USP, Pharmaceutical Associates, Inc., Greenville, SC) daily for five days.

Drug treatment

Rats were given intraperitoneal (*i.p.*) injections of vehicle (VEH: 95% saline + 5% DMSO) or clozapine-N-oxide, CNO (3 mg/kg CNO in 95% saline + 5% DMSO) 30 min prior to beginning the behavioral task, as in previous work from our lab (Hart et al., 2020; Stolyarova et al., 2019). As shown in **Figure 1D**, in the discrimination stage upon reaching 80% correct across two consecutive days, rats performed three additional sessions of discrimination learning to test for

mastery and successful retention. Session 3 was performed after a single i.p. vehicle (VEH) injection, and Sessions 4 and 5 after a single i.p. injection of CNO. If rats did not continue to reach criterion on Session 5 of CNO, subsequent testing sessions were administered with VEH until the criterion was exceeded after which the reversal stage would commence. Injections of VEH or CNO were administered before every reversal session. Once rats reached 50% correct across two consecutive sessions, the other drug was administered for all subsequent reversal sessions until reaching the 80% criterion. Administration of CNO or VEH for the first half (<50%) and VEH or CNO for the second half (>50%) of reversal sessions was counterbalanced across animals.

Stimulus-Outcome learning

After one week of post-surgery recovery, rats began the stimulus-based discrimination learning task (**Figure 1B**). This task was described in detail previously (Aguirre et al., 2020; Harris et al., 2020), however, in the present design we assigned deterministic stimulus-reward assignments – correct (100%) vs incorrect (0%). Rats initiated each trial by nosepoking a white graphic stimulus in the center screen (displayed for 40 s). The disappearance of the initiation stimulus was immediately followed by two distinct visual stimuli ('fan' vs. 'marbles') presented on the left and right side of the touch screen (displayed for 60 s) pseudorandomly on each trial. One stimulus was paired with a sucrose pellet reward 100% of the time while the other stimulus was not rewarded. Trials that were not initiated within 40 s were scored as an *initiation omission*, whereas failure to select a choice stimulus was scored as a *choice omission*. Non-rewarded trials were followed by a 5 s time-out, and all trials including omissions were concluded with a 10 s ITI before commencing the next trial. The overall learning criterion constituted meeting the requirement of consuming 60 or more rewards along with selecting the correct option >80% of the trials within a 60 min session across two consecutive days. As above, upon reaching criterion, rats performed an additional three sessions after receiving an injection of VEH or CNO. Rats would then advance to the reversal phase in which the previously correct stimulus was now the incorrect stimulus. Methods and criterion were identical to those described above. Rats then received injections of VEH or CNO on the first day of reversal, and on reaching 50% correct (similarly across 2 days and with 60 or more rewards collected), drug injections were switched until reaching the 80% criterion. Injections of VEH or CNO during the reversal stage was counterbalanced across animals.

Action-Outcome learning

All rats who successfully completed the S-O phase advanced to an action-outcome (A-O) version of the task (**Figure 1C**). Methods and criteria were identical to the deterministic stimulus-outcome tasks described above, with the only difference being that the two stimuli used were identical, and the rat was now required to ignore the stimuli and instead respond left or right for reward. The assignments of stimuli ('fan' vs. 'fan' or 'marbles' vs. 'marbles') were counterbalanced across animals.

590 *Electrophysiological Recordings*

591 Data were acquired from a multi-channel data acquisition system (Blackrock
592 Microsystems, LLC., Salt Lake City, UT). The digitizing headstage was connected to the electrode
593 array, and signal was sent to the recording system through a tether and commutator (Dragonfly
594 Inc. Ridgeley, WV). Recording sessions were conducted on all behavioral sessions preceded by
595 an injection. Neural recordings began 15 min after drug injection followed by a 15 min baseline
596 while the animal was inside the operant chamber. The behavioral task commenced at the end of
597 baseline, a total of 30 min post-injection. Four behavioral task epochs (e.g., trial initiation,
598 correct/incorrect stimulus choice, and reward port entry) were configured triggered event
599 outputs from each operant chamber to the data acquisition system that were simultaneously
600 collected with the electrophysiological recordings. 'Off task' neural data were also obtained
601 following injections while animals were not on the learning task, but simply placed in the operant
602 chamber. Random -2/+2 sec epochs of the 60 min recording session can be found in ACC and OFC
603 in **Figure S2E and S2F** for S-O and **Figure S3E and S3F** for A-O learning. Recording data for the
604 EGFP control animal for all trial epochs are shown in **Figure S5** and **Figure S6**, for S-O and A-O
605 learning, respectively.
606

607 *Histology*

608 At the conclusion of the experiment, rats were anesthetized and administered electrolytic
609 lesions at the recording sites via direct current stimulation (20 μA for 20 s), **Figure 3B**. Rats were
610 euthanized 3 days later by sodium pentobarbital (Euthasol, 0.8 mL, *i.p.*; Virbac, Fort Worth, TX)
611 and brains were extracted via transcardial perfusion with phosphate buffered saline (PBS)
612 followed by 10% buffered formalin acetate and post-fixed in this solution for 24 hours followed
613 by 30% sucrose cryoprotection. Tissue was prepared in 40-μM thick coronal sections and either
614 Nissl stained for verification of electrode placement or cover slipped with DAPI mounting
615 medium (Prolong gold, Invitrogen, Carlsbad, CA) and amplified with NMDAR1 Polyclonal
616 Antibodies (Invitrogen for Thermo Fischer Scientific) for DREADDs verification, visualized using
617 a BZ-X710 microscope (Keyence, Itasca, IL), and analyzed with BZ-X Viewer software. The spread
618 of DREADDs expression was determined by matching histological sections to a standard rat brain
619 atlas (Paxinos and Watson, 2007), **Figure 3A and 3B**.
620

621 *Spectral power and Coherence*

622 Local field potential (LFP) data were acquired at 30 kHz and downsampled to 1,000 Hz.
623 Traces were analyzed for signal artifacts through visual inspection and automatic artifact
624 detection. Spectral power across frequency bands were determined through a fast Fourier
625 transform via the *spectrogram()* function in MATLAB [frequency bin = 0.5 Hz, 10 s Hanning
626 window]. LFP Coherence (magnitude-squared coherence) was calculated using Welch's
627 averaged modified periodogram method with a 1-s window and a frequency resolution of 1 Hz
628 via MATLAB function *mscohere()*. LFPs were band-pass filtered for theta (5-10 Hz) using the
629 *filtfilt()* function in MATLAB and designed using the *designfilt()* function (stop bands: 2 Hz, 15Hz;
630 pass bands: 5Hz, 10Hz). ACC-OFC coherence for S-O and A-O learning is shown in **Figure 3C**
631 **and 3D**.

Data analysis

All behavioral and neurophysiological analyses were performed via custom-written code in MATLAB (MathWorks, Inc., Natick, MA). We first demarcated 4 unique task triggered events for our analyses: 1) trial initiation (center stimulus) nosepoke, 2) nosepoke to S^+ (correct stimulus or action), 3) nosepoke to S^- (incorrect stimulus or action), and 4) reward collection (food magazine head entry). LFP signals that exceeded absolute value of 1.5 mV or when summed cross-band power (2 – 120 Hz) exceeded the 99.98th percentile, were identified as artifact and excluded from all analyses. Band-passed theta traces were synchronized from -2 to +2 seconds to each of the four behavioral task triggered events for statistical analyses.

Learning and performance (speed/latency) data were analyzed with a series of mixed-effects General Linear Models (GLMs) (*fitglme* function; Statistics and Machine Learning Toolbox) first in omnibus analyses that included all factors and both learning phases (discrimination and reversal). Then, each learning phase was analyzed separately for measures where significant interactions of phase were observed. Drug condition (CNO vs VEH), ACC and OFC theta power were included in the models as fixed factors, and individual rat was included as a random factor. All post-hoc tests were corrected for the number of comparisons. Statistical significance was noted when p-values were less than 0.05, and p-values between 0.05 and 0.08 were reported as marginally significant. For theta power values that were identified as exceeding model prediction in the GLMs, we sought to further assess whether it was the 2-sec pre- and/or 2-sec post- 'time zero' wherein the theta power was significant. We also analyzed correct vs. incorrect theta power in these choice epochs using independent-sample t-tests as a follow-up to GLM results.

Major dependent variables included: percent correct (and correct and incorrect choices, probed separately), rewards collected (sucrose pellets earned), and median latencies (to initiate a trial, to make the correct choice, to make the incorrect choice, and to collect reward). For A-O learning data, because rats learned so quickly, percent correct was binned into 12 blocks containing 5 trials each, and calculated as a sliding average across all twelve blocks with a 3-trial overlap between each block, for a *Sliding Correct* measure. Pearson correlations for latency measures and theta power were generated using the *corrcoef* function in MATLAB. Latency datapoints exceeding 2 SD were excluded. For latency values, time at trial event (t=0) was subtracted from the previous trial event (t-1). For example, correct or incorrect choice speed was calculated by subtracting time of choice from time of trial initiation, and reward collection speed was calculated by subtracting time of reward collection from time of correct or incorrect choice, in seconds (sec).

EXTENDED DATA

Below we report the omnibus analyses that justified separate analyses of learning phases (discrimination and reversal), reported above.

Global measures of S-O learning: Accuracy and Rewards Collected

For S-O learning, there were many phase interactions for *choice of correct stimulus* or *choice of incorrect stimulus*. However, *rewards collected* was the only variable significantly modulated by OFC and ACC theta power at this level.

Accuracy. We fitted GLMs that included data from all sessions of learning across both phases (discrimination, reversal), and following each drug condition (VEH, CNO) on measures that we previously defined as measures of accuracy (Harris et al. (2020): correct, incorrect, and reward collection). For *choice of the correct stimulus,* we found a significant effect of phase (GLM: β_{phase} =-0.486, t(82)=-7.227, p=2.32e^{-10}); as expected, choice of the correct stimulus occurred more frequently in discrimination than reversal (**Figure 2**). We also observed significant phase x drug (GLM: $\beta_{phase\ x\ drug}$ =0.373, t(82)=2.681, p=0.009) and phase x session (GLM: $\beta_{phase\ x\ session}$ =0.040, t(82)=3.459, p<0.001) interactions. There was also a phase x drug x session interaction (GLM: $\beta_{phase\ x\ drug\ x\ session}$ =-0.125, t(82)=-2.108, p=0.038). Session interactions at this level are of limited value because both phases are included in the analysis.

Following an identical analysis for *choice of the incorrect stimulus*, we observed a near-identical pattern of results after fitting GLMs. We found a significant effect of phase (GLM: β_{phase} =0.470, t(82)=6.876, p=1.118e^{-9}); as expected, choice of incorrect stimulus occurred more frequently in reversal than in discrimination (**Figure 2**). We also observed significant phase x drug (GLM: $\beta_{phase\ x\ drug}$ =-0.412, t(82)=-2.941, p=0.004) and phase x session (GLM: $\beta_{phase\ x\ session}$ =-0.040, t(82)=-3.327, p=0.001), and a phase x drug x session interaction (GLM: $\beta_{phase\ x\ drug\ x\ session}$ =0.143, t(82)=2.423 p=0.02).

Rewards collected. Rats increased their number of rewards collected across sessions (GLM: $\beta_{session}$=23.82, t(82)=2.280, p=0.03) and expectedly collected fewer in the reversal phase compared to discrimination (GLM: β_{phase}=-18.902, t(82)=-3.250, p=0.002). Further, ACC theta and OFC theta power were significant predictors of reward collection, higher when collecting reward (GLM: $\beta_{ACCtheta}$ =5936.4, t(82)=2.454, p=0.016; $\beta_{OFCtheta}$ =5047.2, t(82)=2.444, p=0.017). There was a significant phase x drug (GLM: $\beta_{phase\ x\ drug}$ =31.885, t(82)=2.717, p=0.008), phase x session (GLM: $\beta_{phase\ x\ session}$ =2.346, t(82)=2.287, p=0.025), and session x OFC theta (GLM: $\beta_{session\ x\ OFCtheta}$ = -773.22, t(82)=-2.378, p=0.020) interaction. Similar to choice of correct stimulus and incorrect stimulus epochs described above, we also observed significant interactions of phase x drug x session (GLM: $\beta_{phase\ x\ drug\ x\ session}$ =-13.245, t(82)=-2.693, p=0.009). We observed significant interactions of ACC and OFC theta for this trial epoch only (GLM: $\beta_{ACCtheta\ x\ OFCtheta}$=-1.6145e^{+5}, t(82)=-2.538, p=0.013).

Due to the correlated nature of ACC and OFC theta power values (**Figures 3C and 3D**), we conducted an additional GLM model without ACC theta power as a predictor of accuracy variables, and found that, unlike ACC theta, OFC theta power alone did not exceed model prediction of correct and incorrect choice (choice of correct stimulus, GLM: $\beta_{OFCtheta}$=2.399, t(75)=0.418, p=0.677; choice of incorrect stimulus, GLM: $\beta_{OFCtheta}$=-3.851, t(75)=-0.629, p=0.531). It was, however, a significant predictor of reward collection when ACC theta was removed from the model (GLM: $\beta_{OFCtheta}$=1499.2, t(75)=3.261, p=0.002).

To summarize, these omnibus GLMs on accuracy measures across both discrimination and reversal phases revealed consistent phase interactions that justified analyses of each phase separately. Moreover, significant interaction of ACC and OFC theta power predicted reward collection across sessions. Similar to our recent report (Harris et al., 2020), we also find that animals performed more accurately and collected more rewards in the discrimination phase, compared to the reversal phase.

Global measures of A-O learning: Accuracy and Rewards Collected

For A-O learning, there was strong ACC and OFC theta modulation across phases for *choice of correct response*, whereas OFC theta was the most significant predictor of *choice of incorrect response*. Unlike in S-O learning, the *rewards collected* variable was not significantly modulated by OFC or ACC theta.

Accuracy and rewards collected. Just as with S-O data, we fitted GLMs that included all sessions of learning across both phases (discrimination, reversal), and following each drug condition (VEH, CNO) on measures of accuracy. Overall, we found much more evidence of ACC and OFC theta modulation across accuracy measures in A-O learning compared to S-O, but no phase interactions, suggesting a fairly uniform role across discrimination and reversal phases. For *choice of the correct response,* we found a significant effect of drug (GLM: β_{drug} =-30.964, t(8)=-3.759, p=0.006; with CNO decreasing this choice compared to VEH), session (GLM: $\beta_{session}$=-14.854, t(8)=-3.445, p=0.009), ACC theta (GLM: $\beta_{ACCtheta}$=-1210.7, t(8)=-3.713, p=0.006), and OFC theta (GLM: $\beta_{OFCtheta}$=-1462.5, t(8)=-3.8465, p=0.005). There were several significant 2-way interactions: drug x session (GLM: $\beta_{drug\ x\ session}$ =14.588, t(8)=-3.299, p=0.011), drug x ACC theta (GLM: $\beta_{drug\ x\ ACCtheta}$ =1043.6, t(8)=-2.989, p=0.017), session x ACC theta (GLM: $\beta_{session\ x\ ACCtheta}$ =584.17, t(8)=-3.379, p=0.009), drug x OFC theta (GLM: $\beta_{drug\ x\ OFCtheta}$ =1649.7, t(8)=4.094, p=0.003), session x OFC theta (GLM: $\beta_{session\ x\ OFCtheta}$ =726.29, t(8)=-3.721, p=0.005), and ACC x OFC theta (GLM: $\beta_{OFCtheta\ x\ ACCtheta}$ =56591, t(8)=-3.8197, p=0.005). There were also several significant 3-way interactions: drug x session x ACC theta (GLM: $\beta_{drug\ x\ session\ x\ ACCtheta}$ =-480.74, t(8)=-2.552, p=0.03), drug x session x OFC theta (GLM: $\beta_{drug\ x\ session\ x\ OFCtheta}$ =-800.91, t(8)=-3.880, p=0.005), drug x ACC theta x OFC theta (GLM: $\beta_{drug\ x\ ACCtheta\ x\ OFCtheta}$ =-57175, t(8)=-3.837, p=0.005), and session x ACC theta x OFC theta (GLM: $\beta_{session\ x\ ACCtheta\ x\ OFCtheta}$ =-27895, t(8)=-3.661, p=0.006). Finally, there was a significant 4-way interaction: drug x session x ACC theta x OFC theta (GLM: $\beta_{drug\ x\ session\ x\ ACCtheta\ x\ OFCtheta}$ =27415, t(8)=-3.542, p=0.007).

With an identical analysis for *choice of the incorrect response*, we observed a different pattern of results to those directly above after fitting GLMs, indicating more phase-specific involvement of theta power, specifically that of OFC theta. Indeed, we found a significant effect of phase (GLM: β_{phase} =0.672, t(7)=6.602, p=0.0003); as expected, choice of incorrect stimulus occurred more frequently in reversal than in discrimination. We also observed a significant effect of drug (GLM: β_{drug} =12.194, t(7)=2.538, p=0.031) with CNO resulting in more choices of the incorrect stimulus than VEH, a significant effect of OFC theta (GLM: $\beta_{OFC\ theta}$ =458.71, t(7)=2.694, p=0.031) and significant interactions of phase x drug (GLM: $\beta_{phase\ x\ drug}$ =-0.645, t(7)=-4.762, p=0.002) and phase x session (GLM: $\beta_{phase\ x\ session}$ =-0.294, t(7)=-4.949, p=0.002). There were also significant interactions of drug x OFC theta (GLM: $\beta_{drug\ x\ OFCtheta}$ =-644.68, t(7)=-3.461, p=0.011), phase x drug x session (GLM: $\beta_{phase\ x\ drug\ x\ session}$ =0.3054, t(7)=-3.705, p=0.008), and drug x ACC theta x OFC theta (GLM: $\beta_{drug\ x\ ACCtheta\ x\ OFCtheta}$ =20783, t(7)=2.811, p=0.026). For *rewards collected*, unlike for S-O learning we found no evidence of significant phase, group, session, or ACC or OFC theta modulation of this variable in A-O learning.

Due to the correlated nature of ACC and OFC theta power values (**Figures 3C and 3D**), similar to the S-O analysis we also conducted an analysis of A-O learning this time using a model without OFC theta power as a predictor of accuracy variables, and found that ACC theta alone did not exceed model prediction of accuracy variables (choice of correct response, GLM: $\beta_{ACCtheta}$=-

4.888, t(16)=-0.152, p=0.881; choice of incorrect response, GLM: $\beta_{ACCtheta}$=-13.624, t(15)=-0.506, p=0.620).

Since learning proceeded so rapidly in A-O (**Figure 2)**, we ran trial block (*Sliding Correct*) GLM analyses. When fitting a GLM to block data, we found several phase interactions, and effects of ACC theta, OFC theta, and drug. We observed a significant effect of drug (GLM: β_{drug} =471.12, t(303)=2.788, p=0.006), a significant effect of phase (GLM: β_{phase} =-49.16, t(303)=-5.586, p=5.185e^{-08}), ACC theta (GLM: $\beta_{ACC\ theta}$ =21780, t(303)=3.21, p=0.001), OFC theta (GLM: $\beta_{OFC\ theta}$ =18934, t(303)=4.216, p=3.282e^{-05}) and significant interactions of phase x drug (GLM: $\beta_{phase\ x\ drug}$ =47.461, t(303)=-4.167, p=4.043e^{-05}) and phase x session (GLM: $\beta_{phase\ x\ session}$ =12.504, t(303)=2.112, p=0.036) interaction. There were also significant interactions of drug x OFC theta (GLM: $\beta_{drug\ x\ OFCtheta}$ =-13843, t(303)=-4.012, p=0.008), an ACC theta x OFC theta interaction (GLM: $\beta_{ACCtheta\ x\ OFCtheta}$ =-5.861e^{05}, t(303)=-3.485, p<0.0001), and drug x ACC theta x OFC theta (GLM: $\beta_{drug\ x\ ACCtheta\ x\ OFCtheta}$ =3.712e05, t(303)=2.061, p=0.04). Collectively, these phase interactions justified separate analyses of discrimination and reversal learning (reported above).

To summarize, these omnibus GLMs on accuracy measures at both the session and trial block levels revealed involvement of ACC and OFC theta across phases of A-O learning for *choice of correct response*, and a more phase-specific role in reversal of OFC theta modulation for *choice of the incorrect response,* at the session level. Theta power in neither region significantly predicted *rewards collected*, thus we did not follow-up with further analyses for that measure.

Global measures of S-O Speed: Latencies

Omnibus GLMs on latencies across discrimination and reversal phases revealed theta modulation of a narrower set of trial epochs compared to measures of accuracy and rewards collected.

OFC theta modulation of initiation speed. We fitted GLMs that similarly combined analysis of all sessions across both phases (discrimination and reversal), in either drug condition (VEH vs. CNO), on measures of latency or speed (to initiate a trial, select the correct stimulus, select the incorrect stimulus, or collect reward). For *initiation latencies,* we found a significant effect of phase (GLM: β_{phase} =1.082, t(82)=2.341, p=0.022), significant effect of OFC theta power (GLM: $\beta_{OFCtheta}$ =577.12, t(82)=2.807, p=0.006), and a significant interaction of OFC theta x drug during this trial epoch (GLM: $\beta_{OFCtheta\ x\ drug}$ =-757.53, t(82)=-2.204, p=0.03). There were no other significant effects of drug, session, theta, or their interactions.

GLMs for *correct choice latencies* and *incorrect choice latencies* revealed a significant interaction of phase x drug x session for *incorrect choice latencies* only (GLM: $\beta_{phase\ x\ drug\ x\ session}$ =-0.585, t(82)=-3.66, p<0.001). There were no significant effects of phase, drug, session, or theta interactions in this trial epoch. Lastly, a GLM on *reward collection latencies* revealed a significant interaction of phase x drug (GLM: $\beta_{phase\ x\ drug}$ =-1.006, t(82)=-2.684, p=0.009).

In summary, theta power in OFC modulated initiation speed. As there were significant phase interactions only for *incorrect choice latencies* and *reward collection latencies*, we probed these further by separate analyses of discrimination and reversal phases.

Global measures of A-O Speed: Latencies

We fitted GLMs that similarly combined analysis of all sessions across both phases (discrimination and reversal), in either drug condition (VEH vs. CNO), on measures of speed (i.e., initiate a trial, select the correct stimulus, select the incorrect stimulus, or collect reward).

OFC theta modulates trial initiation speed. For *initiation latencies,* we found a significant effect of OFC theta (GLM: $\beta_{OFCtheta}$ =4567, t(8)=2.438, p=0.041), a significant interaction of phase x drug (GLM: $\beta_{phase\ x\ drug}$=-3.354, t(8)=-2.921, p=0.019), and a drug x OFC theta interaction (GLM: $\beta_{OFCtheta\ x\ drug}$=-5266.8, t(8)=-2.698, p=0.027) during this trial epoch. There was also a significant interaction of ACC x OFC theta (GLM: $\beta_{ACCtheta\ x\ OFCtheta}$=-1.715e^{05}, t(8)=-2.373, p=0.045). There were no other significant effects of drug, session, theta, or interactions.

Both ACC and OFC theta modulate correct choice speed. There were significant effects of phase (GLM: β_{phase}=-1.603, t(8)=-3.019, p=0.017; slower in reversal), drug (GLM: β_{drug}=54.176, t(8)=2.950, p=0.018; quicker with CNO), session (GLM: $\beta_{session}$=30.196, t(8)=3.140, p=0.014), ACC theta (GLM: $\beta_{ACCtheta}$=2106.1, t(8)=2.897, p=0.02), and OFC theta (GLM: $\beta_{OFCtheta}$=2598.3, t(8)=3.065, p=0.015). There were significant 2-way interactions: phase x drug (GLM: $\beta_{phase\ x\ drug}$=2.200, t(8)=-3.578, p=0.007), phase x session (GLM: $\beta_{phase\ x\ session}$=0.929, t(8)=3.555, p=0.007), drug x session (GLM: $\beta_{drug\ x\ session}$=-28.286, t(8)=-2.869, p=0.02), drug x ACC theta (GLM: $\beta_{drug\ x\ ACCtheta}$=-1800.5, t(8)=-2.313, p=0.049), session x ACC theta (GLM: $\beta_{session\ x\ ACCtheta}$=-1182.7, t(8)=-3.068, p=0.015), drug x OFC theta (GLM: $\beta_{drug\ x\ OFCtheta}$=-2949.1, t(8)=-3.282, p=0.011), session x OFC theta (GLM: $\beta_{session\ x\ OFCtheta}$=-1433.8, t(8)=-3.294, p=0.011), and ACC x OFC theta (GLM: $\beta_{ACCtheta\ x\ OFCtheta}$=-1.006e^{05}, t(8)=-3.045, p=0.016). There were significant 3-way interactions as well: phase x drug x session (GLM: $\beta_{phase\ x\ drug\ x\ session}$=-1.2137, t(8)=-3.916, p=0.004), drug x session x OFC theta (GLM: $\beta_{drug\ x\ session\ x\ OFCtheta}$=1493.1, t(8)=3.244, p=0.012), drug x ACC theta x OFC theta (GLM: $\beta_{drug\ x\ ACCtheta\ x\ OFCtheta}$=1.004e^{05}, t(8)=3.021, p=0.017), and session x ACC theta x OFC theta (GLM: $\beta_{session\ x\ ACCtheta\ x\ OFCtheta}$=54649, t(8)=3.217, p=0.012). Finally, there was a significant 4-way interaction of drug x session x OFC theta x ACC theta ($\beta_{drug\ x\ session\ x\ ACCtheta\ x\ OFCtheta}$=-50909, t(8)=-2.950, p=0.018).

ACC and OFC theta do not modulate incorrect choice or reward collection speed. GLM analyses for these variables revealed no significant role of ACC and OFC theta to either incorrect choice or reward collection speed. In summary, similar to S-O learning, theta power in OFC modulated initiation speed in A-O. As there were significant phase interactions only for *correct choice latencies*, we probed these further by separate analyses of discrimination and reversal phases, above.

Acknowledgements: This work was supported by UCLA's Division of Life Sciences Retention fund (Izquierdo), R01 DA047870 (Izquierdo and Soltani), R21 MH122800 (Izquierdo and Blair), the Training program in Neurotechnology Translation T32 NS115753 (Ye). We appreciate early and helpful comments from members of the Izquierdo lab on these data. 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.

Conflict of Interest: The authors declare no competing financial interests

843 ***Table 1. ACC and OFC theta modulation of accuracy and speed in S-O and A-O learning.***

	Accuracy	
	Choice of Correct	Choice of Incorrect
S-O		
Discrimination	ACC	ACC x OFC
Reversal	OFC	OFC
A-O		
Discrimination	ACC x OFC	n/a
Reversal	OFC	n/a
	Speed	
	Choice of Correct	Choice of Incorrect
S-O		
Discrimination	X	X
Reversal	X	X
A-O		
Discrimination	X	X
Reversal	ACC x OFC	X

844 Summary of mixed-effects GLMs post-hoc analyses of major trial epochs. X indicates that ACC and/or OFC
845 theta power were not significant predictors of those trial epochs. Not shown here are Trial Initiation and
846 Reward Collection epochs, for which there were no significant post-hoc effects. n/a = A *Sliding Correct*
847 measure was obtained due to the quicker learning in A-O compared to S-O.
848
849
850
851
852
853
854
855
856
857
858
859

Supplementary Figures

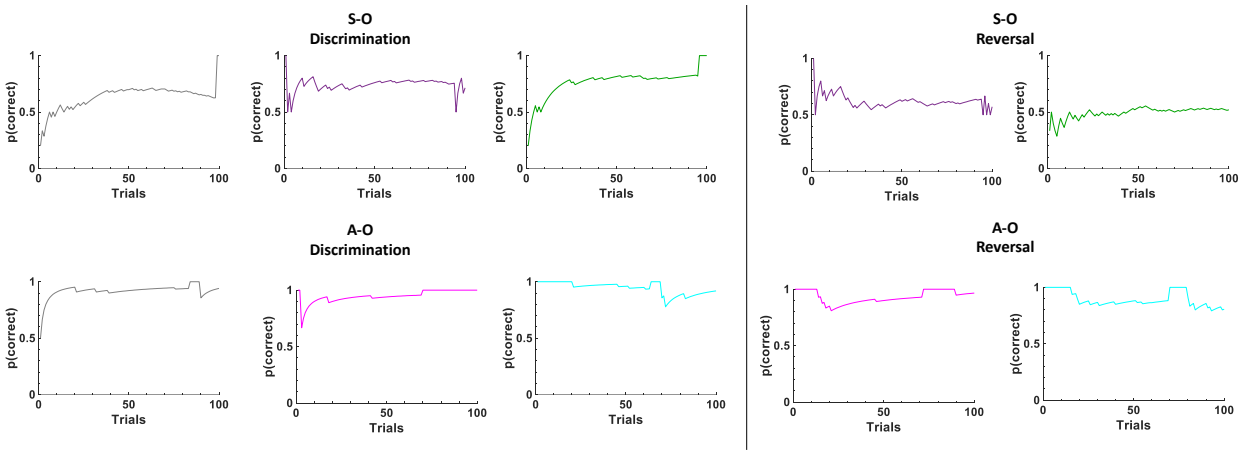


Figure S1. Trial-by-trial performance within a single session during different learning stages. Rats learned A-O associations much more quickly than S-O associations, the latter often within a single session. Shown is representative trial-by-trial performance of the first 100 trials within different learning stages. Learning during S-O initial discrimination and reversal for no recording days (No Rec, grey), and sessions preceded by clozapine-N-oxide (CNO, purple) or vehicle (VEH, green) injections (top rows). Learning during A-O initial discrimination and reversal for no recording days (No Rec, grey), and sessions preceded by clozapine-N-oxide (CNO, magenta) or vehicle (VEH, cyan) injections (bottom rows).

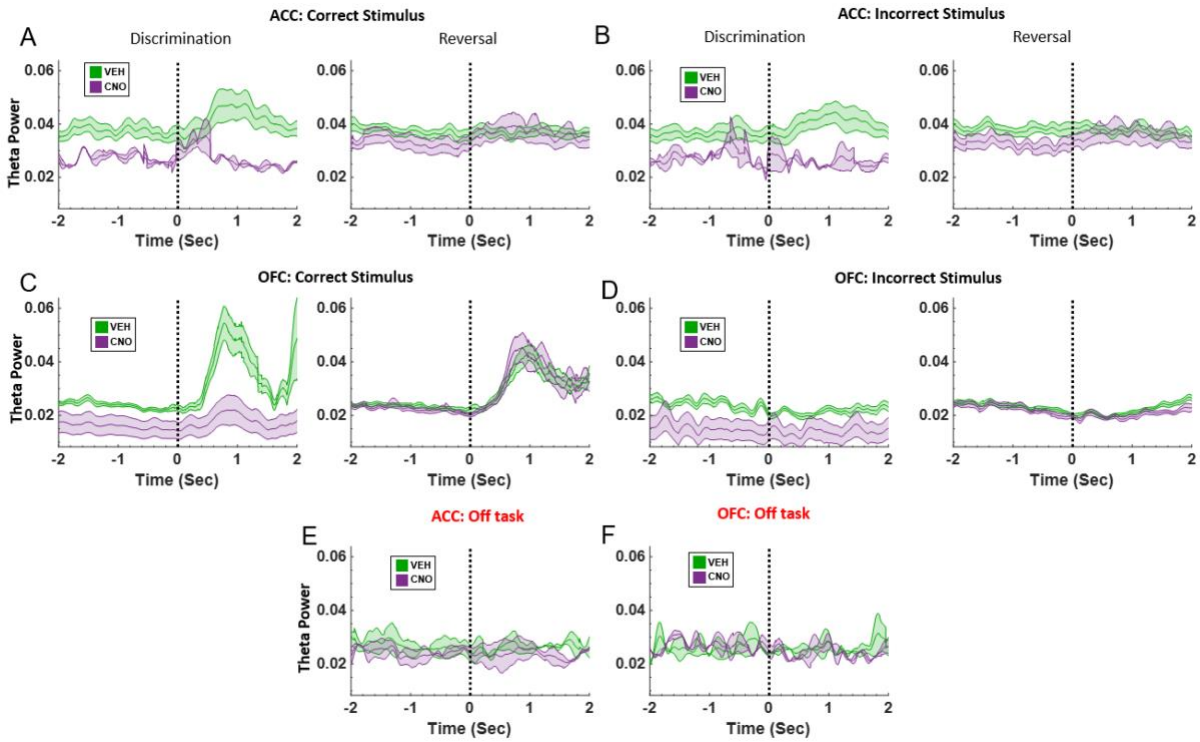


Figure S2. Theta power during significant trial epochs in S-O learning and "Off task" conditions. Animals that were prepared with inhibitory Gi DREADDs in ACC were assessed for accuracy in S-O learning that included choice of correct stimulus and choice of incorrect stimulus. **(A-B)** ACC theta power during choice of the correct stimulus and choice of incorrect stimulus during discrimination learning (left) and reversal learning (right) phases. (**C-D)** OFC theta power during choice of the correct stimulus and choice of incorrect stimulus during discrimination learning (left) and reversal learning (right) phases. Animals were injected with clozapine-N-oxide (CNO, purple) or vehicle (VEH) 30 min before testing (**E-F)** Random -2/+2 sec epochs throughout the 60 min recording sessions were averaged while animals with DREADDs in ACC were 'off task.' Animals were injected with clozapine-N-oxide (CNO, purple) or vehicle (VEH) 30 min before this "off task" session was conducted. Theta power mean $\pm$ standard error of the mean (S.E.M.) shading.

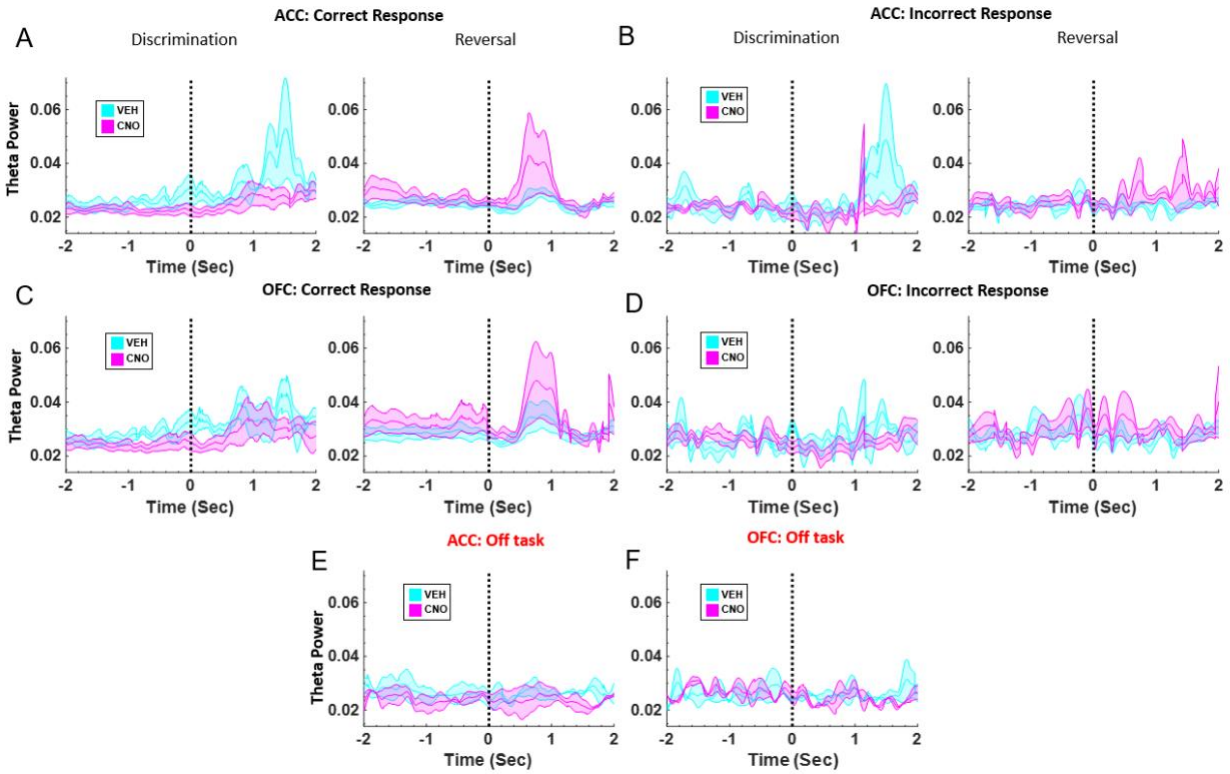


Figure S3. Theta power during significant trial epochs in A-O learning and "Off task" conditions. Animals that were prepared with inhibitory Gi DREADDs in ACC were assessed for accuracy in A-O learning that included choice of correct response and choice of incorrect response. **(A-B)** ACC theta power during choice of the correct response and choice of incorrect response during discrimination learning (left) and reversal learning (right) phases. (**C-D)** OFC theta power during choice of the correct response and choice of incorrect response during discrimination learning (left) and reversal learning (right) phases. Animals were injected with clozapine-N-oxide (CNO, magenta) or vehicle (VEH, cyan) 30 min before testing (**E-F)** Random -2/+2 sec epochs throughout the 60 min recording sessions were averaged while animals with DREADDs in ACC were 'off task.' Animals were injected with clozapine-N-oxide (CNO, magenta) or vehicle (VEH, cyan) 30 min before this "off task" session was conducted. Theta power mean $\pm$ standard error of the mean (S.E.M.) shading.

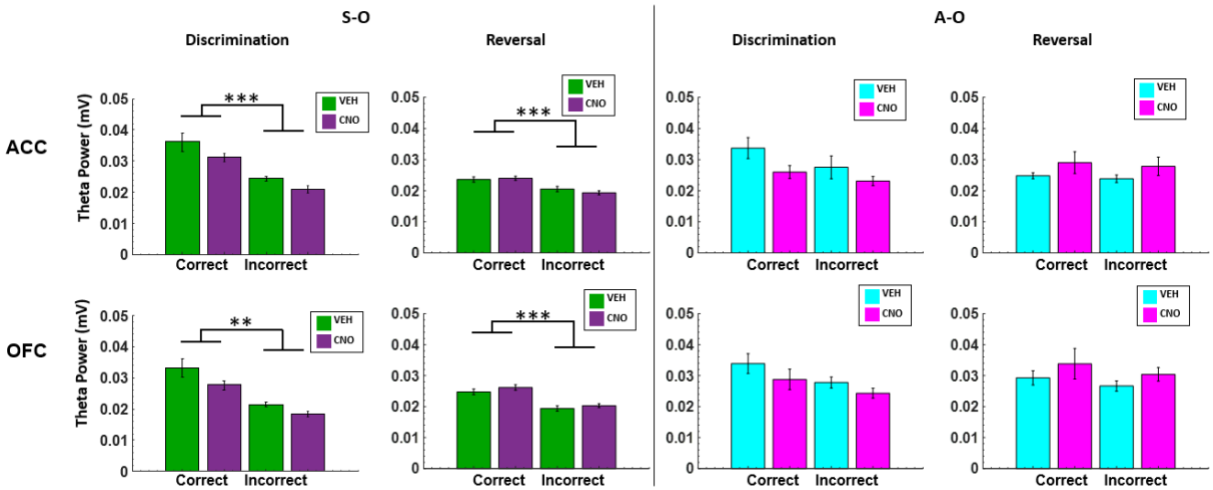
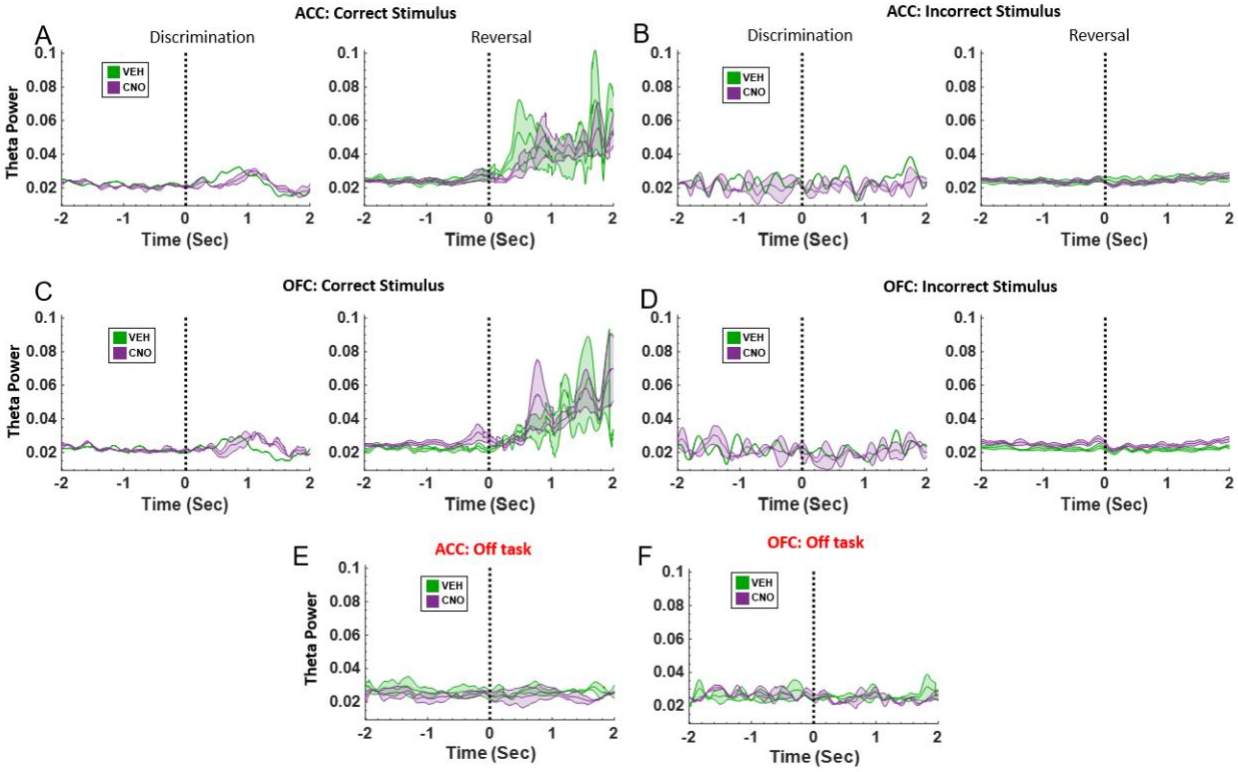


Figure S4. Theta power differences in S-O learning accuracy. Theta power was significantly higher in correct choices versus incorrect choices in both ACC (discrimination + reversal) and OFC (discrimination + reversal) during the S-O task. There was no significant difference in theta power between correct choices versus incorrect choices in either ACC or OFC during the A-O task, for either phase. There were no instances of incorrect trials following CNO in the discrimination phase, accounting for the missing bar. Given there was no statistically significant effect of drug in predicting accuracy, both CNO and VEH conditions were averaged together to compare correct vs. incorrect choices. Independent samples t-tests, bars indicate $\pm$ standard error of the mean (S.E.M.) **p<0.01, ***p<0.001.

887



888

Figure S5. Theta power during significant trial epochs in S-O learning for eGFP null virus control. One animal was prepared with eGFP null virus in ACC and assessed for accuracy in S-O learning that included choice of correct stimulus and choice of incorrect stimulus. **(A-B)** ACC theta power during choice of the correct stimulus and choice of incorrect stimulus during discrimination learning (left) and reversal learning (right) phases. (**C-D)** OFC theta power during choice of the correct stimulus and choice of incorrect stimulus during discrimination learning (left) and reversal learning (right) phases. The animal was injected with clozapine-N-oxide (CNO, purple) or vehicle (VEH) 30 min before testing (**E-F)** Random -2/+2 sec epochs throughout the 60 min recording sessions were averaged while the animal with eGFP null virus in ACC was 'off task.' The animal was injected with clozapine-N-oxide (CNO, purple) or vehicle (VEH) 30 min before this "off task" session was conducted. Theta power mean $\pm$ standard error of the mean (S.E.M.) shading.

889

890

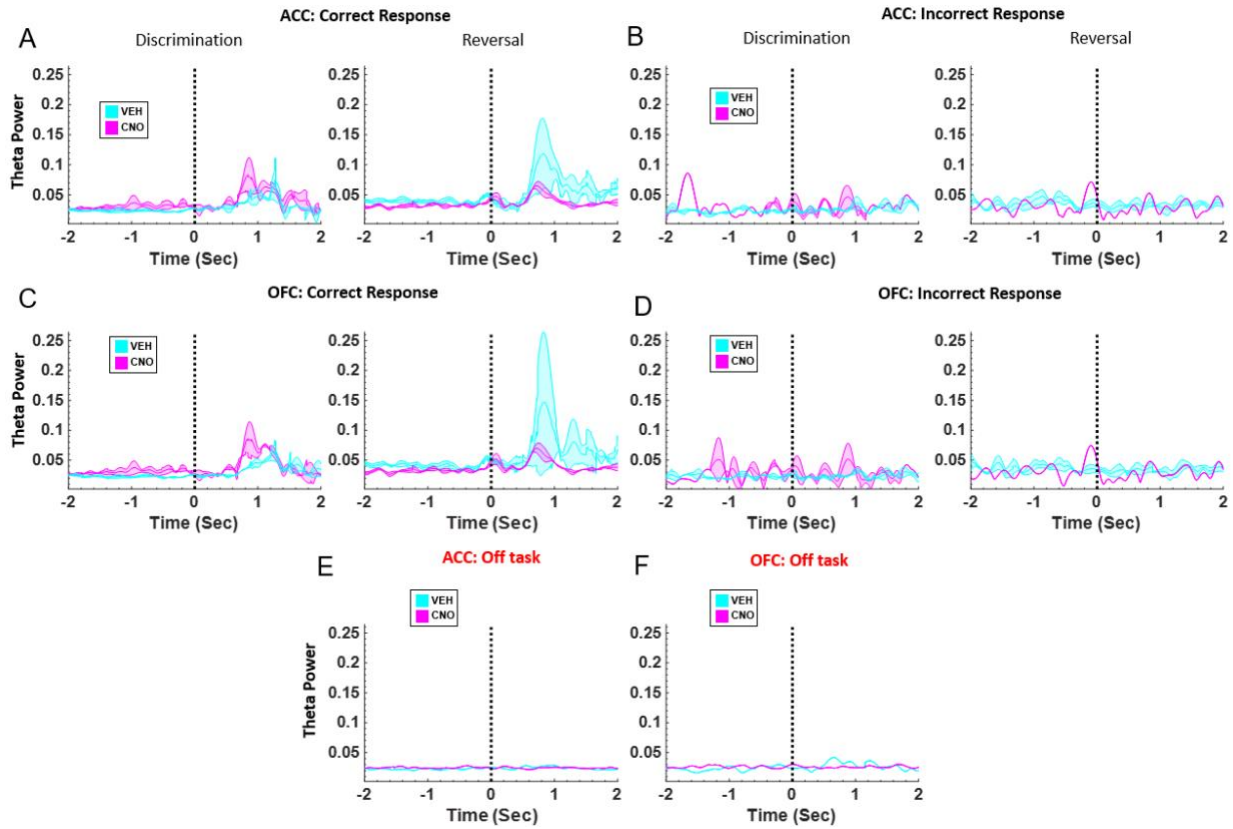


Figure S6. Theta power during significant trial epochs in A-O learning for eGFP null virus control. One animal was prepared with eGFP null virus in ACC and assessed for accuracy in A-O learning that included choice of correct response and choice of incorrect response. (A-B) ACC theta power during choice of the correct response and choice of incorrect response during discrimination learning (left) and reversal learning (right) phases. (C-D) OFC theta power during choice of the correct response and choice of incorrect response during discrimination learning (left) and reversal learning (right) phases. The animal was injected with clozapine-N-oxide (CNO, magenta) or vehicle (VEH, cyan) 30 min before testing (E-F) Random -2/+2 sec epochs throughout the 60 min recording sessions were averaged while the animal with eGFP null virus in ACC was 'off task.' The animal was injected with clozapine-N-oxide (CNO, magenta) or vehicle (VEH, cyan) 30 min before this "off task" session was conducted. Theta power mean $\pm$ standard error of the mean (S.E.M.) shading.

891

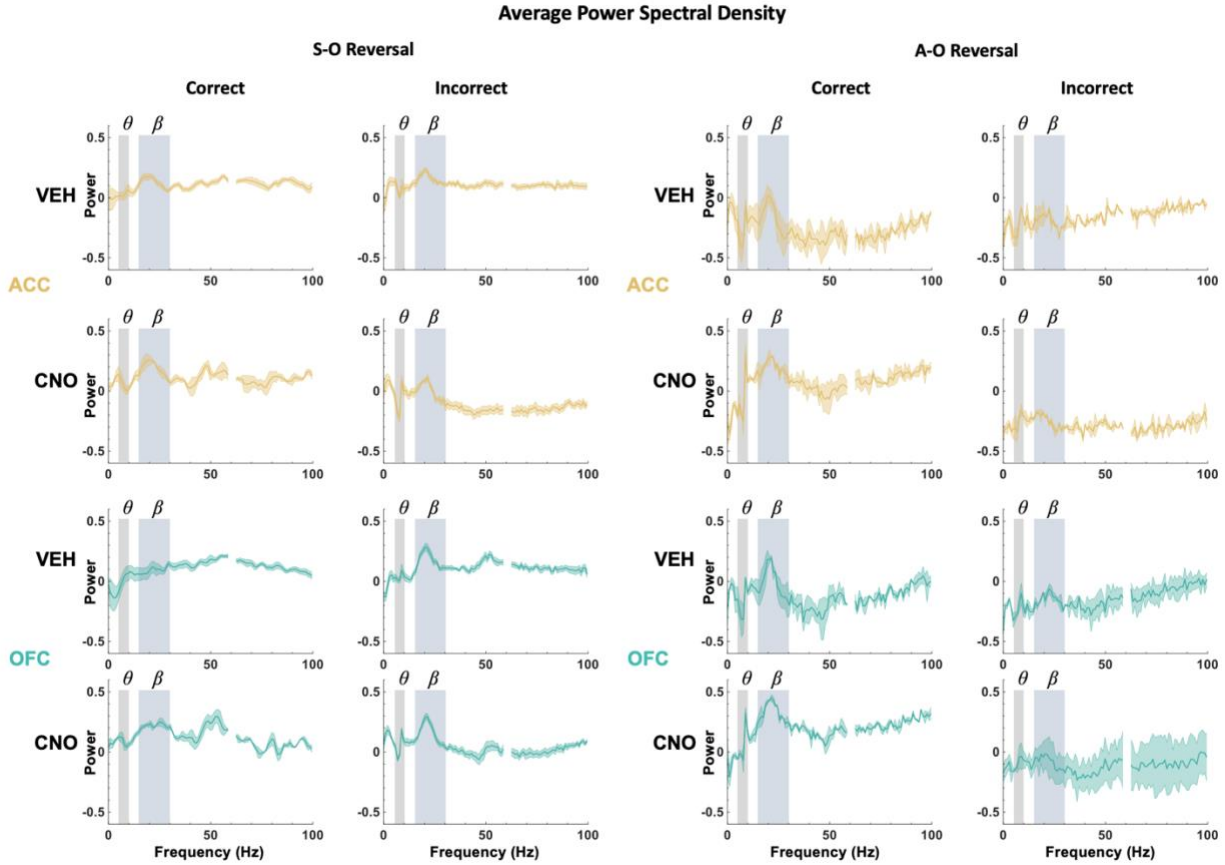


Figure S7. Event-triggered average power spectral density. Spectral power for ACC and OFC was averaged across all ACC DREADDs animals based on task phase (S-O, A-O) and drug (VEH, CNO). Grey shading highlights theta (5 – 10 Hz) and beta (15 – 30 Hz) frequency bands.

892

REFERENCES

Aguirre, C.G., Stolyarova, A., Das, K., Kolli, S., Marty, V., Ray, L., Spigelman, I., and Izquierdo, A. (2020). Sex-dependent effects of chronic intermittent voluntary alcohol consumption on attentional, not motivational, measures during probabilistic learning and reversal. PLoS One *15*, e0234729.

Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R.F., Dayan, P., and Costa, R.M. (2021). The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection. Neuron *109*, 149-163 e147.

Alexander, G.M., Brown, L.Y., Farris, S., Lustberg, D., Pantazis, C., Gloss, B., Plummer, N.W., Jensen, P., and Dudek, S.M. (2018). CA2 neuronal activity controls hippocampal low gamma and ripple oscillations. Elife *7*.

Amarante, L.M., and Laubach, M. (2021). Coherent theta activity in the medial and orbital frontal cortices encodes reward value. eLife *10:e63372*.

Bari, B.A., Grossman, C.D., Lubin, E.E., Rajagopalan, A.E., Cressy, J.I., and Cohen, J.Y. (2019). Stable Representations of Decision Variables for Flexible Behavior. Neuron *103*, 922-933.e927.

Barreiros, I.V., Panayi, M.C., and Walton, M.E. (2021). Organization of Afferents along the Anterior-posterior and Medial-lateral Axes of the Rat Orbitofrontal Cortex. Neuroscience *460*, 53-68.

Behrens, T.E., Woolrich, M.W., Walton, M.E., and Rushworth, M.F. (2007). Learning the value of information in an uncertain world. Nat Neurosci *10*, 1214-1221.

Boulougouris, V., Dalley, J.W., and Robbins, T.W. (2007). Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. Behav Brain Res *179*, 219-228.

Bryden, D.W., Johnson, E.E., Tobia, S.C., Kashtelyan, V., and Roesch, M.R. (2011). Attention for learning signals in anterior cingulate cortex. J Neurosci *31*, 18266-18274.

Bussey, T.J., Muir, J.L., Everitt, B.J., and Robbins, T.W. (1997). Triple dissociation of anterior cingulate, posterior cingulate, and medial frontal cortices on visual discrimination tasks using a touchscreen testing procedure for the rat. Behav Neurosci *111*, 920-936.

Camille, N., Tsuchida, A., and Fellows, L.K. (2011). Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. J Neurosci *31*, 15048-15052.

Chen, C.S., Ebitz, R.B., Bindas, S.R., Redish, A.D., Hayden, B.Y., and Grissom, N.M. (2021). Divergent Strategies for Learning in Males and Females. Curr Biol *31*, 39-50 e34.

Chudasama, Y., Daniels, T.E., Gorrin, D.P., Rhodes, S.E., Rudebeck, P.H., and Murray, E.A. (2013). The role of the anterior cingulate cortex in choices based on reward value and reward contingency. Cereb Cortex *23*, 2884-2898.

Chudasama, Y., Passetti, F., Rhodes, S.E., Lopian, D., Desai, A., and Robbins, T.W. (2003). Dissociable aspects of performance on the 5-choice serial reaction time task following lesions of the dorsal anterior cingulate, infralimbic and orbitofrontal cortex in the rat: differential effects on selectivity, impulsivity and compulsivity. Behav Brain Res *146*, 105-119.

Chudasama, Y., and Robbins, T.W. (2003). Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. J Neurosci *23*, 8771-8780.

Cowen, S.L., Davis, G.A., and Nitz, D.A. (2012). Anterior cingulate neurons in the rat map anticipated effort and reward to their associated action sequences. J Neurophysiol *107*, 2393-2407.

Dalton, G.L., Wang, N.Y., Phillips, A.G., and Floresco, S.B. (2016). Multifaceted Contributions by Different Regions of the Orbitofrontal and Medial Prefrontal Cortex to Probabilistic Reversal Learning. J Neurosci *36*, 1996-2006.

Faraut, M.C., Procyk, E., and Wilson, C.R. (2016). Learning to learn about uncertain feedback. Learn Mem *23*, 90-98.

Fatahi, Z., Ghorbani, A., Ismail Zibaii, M., and Haghparast, A. (2020). Neural synchronization between the anterior cingulate and orbitofrontal cortices during effort-based decision making. Neurobiol Learn Mem *175*, 107320.

Fatahi, Z., Haghparast, A., Khani, A., and Kermani, M. (2018). Functional connectivity between anterior cingulate cortex and orbitofrontal cortex during value-based decision making. Neurobiol Learn Mem *147*, 74-78.

Floresco, S.B., Block, A.E., and Tse, M.T. (2008). Inactivation of the medial prefrontal cortex of the rat impairs strategy set-shifting, but not reversal learning, using a novel, automated procedure. Behav Brain Res *190*, 85-96.

Groman, S.M., Keistler, C., Keip, A.J., Hammarlund, E., DiLeone, R.J., Pittenger, C., Lee, D., and Taylor, J.R. (2019). Orbitofrontal Circuits Control Multiple Reinforcement-Learning Processes. Neuron *103*, 734-746 e733.

Harris, C., Aguirre, C.G., Kolli, S., Das, K., Izquierdo, A., and Soltani, A. (2020). Unique features of stimulus-based probabilistic reversal learning. bioRxiv *2020.09.24.310771*.

Hart, E.E., Blair, G.J., O'Dell, T.J., Blair, H.T., and 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.

Hayden, B.Y., Heilbronner, S.R., Pearson, J.M., and Platt, M.L. (2011). Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J Neurosci *31*, 4178-4187.

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

Hoover, W.B., and Vertes, R.P. (2011). Projections of the medial orbital and ventral orbital cortex in the rat. J Comp Neurol *519*, 3766-3801.

Hunt, L.T., and Hayden, B.Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. Nat Rev Neurosci *18*, 172-182.

Hunt, L.T., Malalasekera, W.M.N., de Berker, A.O., Miranda, B., Farmer, S.F., Behrens, T.E.J., and Kennerley, S.W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. Nat Neurosci *21*, 1471-1481.

Hyman, J.M., Holroyd, C.B., and Seamans, J.K. (2017). A Novel Neural Prediction Error Found in Anterior Cingulate Cortex Ensembles. Neuron *95*, 447-456 e443.

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

Jackson, S.A., Horst, N.K., Pears, A., Robbins, T.W., and Roberts, A.C. (2016). Role of the Perigenual Anterior Cingulate and Orbitofrontal Cortex in Contingency Learning in the Marmoset. Cereb Cortex *26*, 3273-3284.

Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., and Rushworth, M.F. (2006). Optimal decision making and the anterior cingulate cortex. Nat Neurosci *9*, 940-947.

Knudsen, E.B., and Wallis, J.D. (2020). Closed-Loop Theta Stimulation in the Orbitofrontal Cortex Prevents Reward-Based Learning. Neuron *106*, 537-547 e534.

Liu, Y., McAfee, S.S., and Heck, D.H. (2017). Hippocampal sharp-wave ripples in awake mice are entrained by respiration. Sci Rep *7*, 8950.

Luk, C.H., and Wallis, J.D. (2013). Choice coding in frontal cortex during stimulus-guided or action-guided decision-making. J Neurosci *33*, 1864-1871.

MacLaren, D.A., Browne, R.W., Shaw, J.K., Krishnan Radhakrishnan, S., Khare, P., Espana, R.A., and Clark, S.D. (2016). Clozapine N-Oxide Administration Produces Behavioral Effects in Long-Evans Rats: Implications for Designing DREADD Experiments. eNeuro *3*.

Marquardt, K., Sigdel, R., and Brigman, J.L. (2017). Touch-screen visual reversal learning is mediated by value encoding and signal propagation in the orbitofrontal cortex. Neurobiol Learn Mem *139*, 179-188.

Mashhoori, A., Hashemnia, S., McNaughton, B.L., Euston, D.R., and Gruber, A.J. (2018). Rat anterior cingulate cortex recalls features of remote reward locations after disfavoured reinforcements. Elife *7*.

Paxinos, G., and Watson, C. (2007). The rat brain in stereotaxic coordinates, 6th edn (Amsterdam ; Boston ;: Academic Press/Elsevier).

Riceberg, J.S., and Shapiro, M.L. (2012). Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior. J Neurosci *32*, 16402-16409.

Rudebeck, P.H., Behrens, T.E., Kennerley, S.W., Baxter, M.G., Buckley, M.J., Walton, M.E., and Rushworth, M.F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. J Neurosci *28*, 13775-13785.

Rudebeck, P.H., and Izquierdo, A. (2021). Foraging with the frontal cortex: A cross-species evaluation of reward-guided behavior. Neuropsychopharmacology.

Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M., and Rushworth, M.F. (2006). Separate neural pathways process different decision costs. Nat Neurosci *9*, 1161-1168.

Schmidt, B., Duin, A.A., and Redish, A.D. (2019). Disrupting the medial prefrontal cortex alters hippocampal sequences during deliberative decision making. J Neurophysiol *121*, 1981-2000.

Schoenbaum, G., Chiba, A.A., and Gallagher, M. (1999). Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. J Neurosci *19*, 1876-1884.

Schweimer, J., and Hauber, W. (2006). Dopamine D1 receptors in the anterior cingulate cortex regulate effort-based decision making. Learn Mem *13*, 777-782.

Stolyarova, A., and Izquierdo, A. (2017). Complementary contributions of basolateral amygdala and orbitofrontal cortex to value learning under uncertainty. Elife *6*.

Stolyarova, A., Rakhshan, M., Hart, E.E., O'Dell, T.J., Peters, M.A.K., Lau, H., Soltani, A., and Izquierdo, A. (2019). Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. Nat Commun *10*, 4704.

Totah, N.K., Kim, Y.B., Homayoun, H., and Moghaddam, B. (2009). Anterior cingulate neurons represent errors and preparatory attention within the same behavioral sequence. J Neurosci *29*, 6418-6426.

van Heukelum, S., Mars, R.B., Guthrie, M., Buitelaar, J.K., Beckmann, C.F., Tiesinga, P.H.E., Vogt, B.A., Glennon, J.C., and Havenith, M.N. (2020). Where is Cingulate Cortex? A Cross-Species View. Trends Neurosci *43*, 285-299.

Voloh, B., Valiante, T.A., Everling, S., and Womelsdorf, T. (2015). Theta-gamma coordination between anterior cingulate and prefrontal cortex indexes correct attention shifts. Proc Natl Acad Sci U S A *112*, 8457-8462.