Prefrontal dynamics and encoding of flexible rule switching

Marco Nigro[1], Lucas Silva Tortorelli [1], Kevin Dinh[1], Machhindra Garad[1], Natalie E. Zlebnik[2,3], and Hongdian Yang [1,3]

[1] Department of Molecular, Cell and Systems Biology, [2] Division of Biomedical Sciences, School of Medicine, [3] Neuroscience Graduate Program, University of California, Riverside, CA 92521, USA

Correspondence: hongdian@ucr.edu

Abstract

Behavioral flexibility, the ability to adjust behavioral strategies in response to changing environmental contingencies and internal demands, is fundamental to cognitive functions. Despite a large body of pharmacology and lesion studies, the underlying neurophysiological correlates and mechanisms that support flexible rule switching remain elusive. To address this question, we trained mice to distinguish complex sensory cues comprising different perceptual dimensions (set shifting). Endoscopic calcium imaging revealed that medial prefrontal cortex (mPFC) neurons represented multiple task-related events and exhibited pronounced dynamic changes during rule switching. Notably, prominent encoding capacity in the mPFC was associated with switching across, but not within perceptual dimensions. We then showed the involvement of the ascending modulatory input from the locus coeruleus (LC), as inhibiting the LC impaired rule switching behavior and impeded mPFC dynamic processes and encoding. Our results highlight the pivotal role of the mPFC in set shifting processes and demonstrate the profound impact of ascending neuromodulation on shaping prefrontal neural dynamics and behavioral flexibility.

Introduction

The ability to adjust behavioral strategies in response to changing external contexts and internal needs, termed behavioral/cognitive flexibility, requires adaptive processing of environmental cues and internal states to guide goal-oriented behavior, and is vital to the survival of organisms. Inappropriate behavioral adjustments, such as deficits in modifying responses to a rule change, are hallmarks of impaired executive functions and are observed in a broad spectrum of psychiatric disorders [1,2].

Decades of research have strived to uncover the neural substrates of behavioral flexibility (e.g., see reviews [1–6]). Set shifting, a type of rule switching that requires attending to or ignoring a stimulus feature in a context-dependent way, is commonly used to assess flexibility. The Wisconsin Card Sorting Test, the Intra-Extra Dimensional Set Shift Task and their analogs have been widely used to test human and animal subjects [7–14]. Prior research using lesion and pharmacology approaches has provided compelling evidence that the prefrontal cortex (PFC) plays a central role in flexible switching behavior (e.g., [5,7,8,15–18]). However, the circuit mechanisms underlying how the PFC is involved in switching-related processes to support set shifting is poorly understood.

To elucidate the neural substrates underlying behavioral flexibility, we first recorded calcium activity from populations of medial PFC (mPFC) neurons in mice performing a set-shifting task, where they learned to discriminate complex sensory cues comprising different perceptual dimensions (digging medium and odor). We identified subgroups of mPFC neurons representing different task-related variables. We employed analytical tools to describe the dynamic processes of mPFC population activity and found that prefrontal population dynamics exhibited a marked shift associated with rule switching. Interestingly, pronounced mPFC encoding capacity was linked to switching across, but not within perceptual dimensions. We further explored the influence of ascending neuromodulatory input on behavior and PFC activity. Through a combination of chemogenetic inhibition and calcium imaging, we showed that the locus coeruleus (LC) was important for set shifting, such that LC inhibition impaired task performance, affected single-cell representation, and impeded population dynamic processes and encoding. Inhibiting LC terminals in the mPFC similarly impaired behavior, demonstrating the importance of direct LC input to the mPFC. Together, our data reveal a pivotal role

of the mPFC in supporting set-shifting processes and highlight the profound impact of ascending neuromodulation on shaping mPFC dynamics during rule switching.

Results

We trained mice to perform the freely-moving attentional set-shifting task (AST) [19,20] based on procedures described in previous studies [21,22] (Methods). In brief, mice learned to discriminate complex sensory cues by associating a relevant stimulus feature to reward (Fig. 1a, b). In most stages of the task (simple discrimination, SD; compound discrimination, CD; intra-dimensional reversal, REV; intra-dimensional shift, IDS), the relevant feature remained in the perceptual dimension of digging medium. In the stage of extra-dimensional shift (EDS), the relevant feature shifted to the dimension of odor. Mice learned each rule change in a single session, but typically took more trials to complete REV and EDS (Fig. 1c) [19,22–24]. Notably, solving EDS requires the recognition of perceptual dimensions (digging medium vs. odor), and implements a reversal component regarding the dimensions instead of any specific cues. This process distinguishes AST from several other set-shifting tasks that do not require the generalization of perceptual dimensions (e.g., [25,26]). It is commonly considered that the ability to switch in EDS reflects the cognitive capacity to generalize and discriminate abstract contextual rules [19,20,27–29], and our current study is focused on revealing the mechanisms underlying this type of rule switching processes.

Single neurons encode various task-related variables. Prior lesion studies provided compelling evidence that the mPFC is involved in EDS switching [15,19,20,30,31]. However, the underlying neurophysiological mechanisms are rarely explored. We expressed calcium indictor GCaMP6f in mPFC and implanted a gradient-index (GRIN) lens. With a head-mounted miniaturized microscope, we monitored calcium activity from populations of mPFC neurons at single-cell resolution in task-performing mice (Fig. 1d, Methods. 368 neurons from 4 mice). We first classified the representation of individual mPFC neurons based on their activity before decision (onset of digging). Supporting a series of previous work (e.g., [32–38]), we identified subgroups of mPFC neurons responding to different task-related variables, such as choice, trial history and task rule (Fig. 1e, Methods). We further noticed that a sizable amount of these neurons responded to more than one task-related event (> 20%, Supp. Fig. 1) [32,35,39–41].

Pronounced dynamic shifts in population activity reflect rule switching. While describing single-neuron representation provides important insights into neuronal functions, the rest of this paper primarily focused on population-level analysis. This is because ensembles, instead of individual neurons, are proposed to be the functional unit of the nervous system [42–45]. They represent information that is hardly accessible from single neurons [36,46,47], especially in higher-order structures where single neurons exhibit mixed tuning [39,48]. To better understand the involvement of the mPFC in set shifting, we first employed a dimensionality-reduction approach to assess population neural dynamics (Methods). Specifically, we examined whether mPFC dynamic processes represent rule-related information. Since the rule change was not cued, at the beginning of EDS animals followed the previous rule that the medium dimension was relevant [19]. Through continuous trial-and-error learning animals eventually adapted their behavior to follow the new rule [32,34]. Thus, we infer the initial mixture of correct and incorrect trials and the last set of consecutive correct trials as different states of rule switching behavior (Early: trial-and-error learning; Late: rule-following. Fig. 1f). Principle component analysis (PCA) was applied to mPFC population activity measured in the two states, and the separation between the low-dimensional vectors was quantified in Euclidean distance. The same analysis was applied to label-shuffled trials as controls (Methods). We identified an overall prominent trajectory separation between the vectors representing rule-learning and rule-following states (Fig. 1g, h. Euclidean distance during the 5-s window prior to decision, data vs. shuffle: 7.5 ± 0.05 vs. 0.3 ± 0.02, $P < 1e\text{-}3$), showing that a shift in population activity space is associated with the acquisition of the new rule. This result suggests that mPFC dynamics reflect the change of task rule. The trajectory separation also exhibited transient fluctuations prior to decision, first increasing and then decreasing (Fig. 1h), suggesting a dynamic decision-related population encoding process underlying the behavioral transitions.

Next, we employed decoding approaches to explore to what extent mPFC activity can quantitatively distinguish animal behavior during rule switching. We first implemented a hidden Markov model (HMM) to examine whether mPFC activity can predict the inferred states and state transitions in individual sessions [38] (Fig. 2a, Methods). HMM assumes that the underlying cognitive states influence the observed neural activity. However, as these states are not directly observable, they are referred to as hidden states (rule-learning and rule-following states in our case). Using the observed neural activity patterns, the algorithm infers the most probable sequence of hidden states that generate the data. This sequence of hidden states provides insights into the predicted behavioral patterns. We used two measures to quantify HMM performance: 1) the accuracy of predicting the state of individual trials (early or late); and 2) the accuracy of predicting where state transition occurred (the onset of late state: switch point). We first used randomized trials as internal controls. Both measures showed that mPFC activity can robustly predict the behavioral states above the shuffled data (Fig. 2b, state prediction accuracy, data vs. shuffle: 77 ± 2% vs. 61 ± 3%, P < 1e-3; Δ switch point, data vs. shuffle: -2 ± 1 trials vs. -6 ± 1 trials, P = 5.7e-3). We further contrasted model performance in EDS to REV, another challenging stage in AST but the rule change did not cross perceptual dimensions. Interestingly, we found a dramatic reduction of model performance in REV (Fig. 2b-e, state prediction accuracy, EDS vs. REV: 77 ± 2% vs. 65 ± 3%, P = 2e-3; Δ switch point, EDS vs. REV: -2 ± 1 trials vs. -6 ± 1 trials, P < 1e-3). This reduction was not due to a lack of mPFC activity in REV as calcium activity was not different between the two stages (Supp. Fig. 2). Our data provide insights into mPFC's encoding properties during set-shifting behavior, and shed light on the long-standing findings that mPFC loss-of-function specifically affected EDS-type switching (e.g., [15,19,20,31]).

We further wondered whether mPFC activity can track animals' decisions on a trial-by-trial basis. We applied a generalized liner model (GLM) to predict the upcoming choice in individual trials (correct vs. incorrect response). Specifically, we included the first 3 PCs of population activity quantified prior to decision as regressors (Methods). We found that mPFC activity can more precisely predict trial-by-trial choices in EDS than in REV (Fig. 2f, g, EDS vs. REV: 72 ± 1% vs. 60 ± 2%, P < 1e-3). Collectively, these findings provide new insights into mPFC's involvement in AST.

LC inhibition impairs switching behavior. Flexible behavior involves complex cognitive processes, which likely engage other brain areas interacting with the PFC [2,49–53]. Prior lesion and perturbation studies suggest a role of locus coeruleus (LC) neurons as well as their input to the PFC in flexible switching behavior, including AST [24,25,52–57]. However, our understanding of the neurophysiological substrates is limited. We employed a transgenic approach to conditionally express the inhibitory DREADD receptors in genetically-defined LC neurons (AAV-DIO-hM4Di injected in DBH-Cre mice. DBH: dopamine beta hydroxylase. Fig. 3a). LC perturbation was reported to affect other stages of AST [54,56]. To avoid the potential complication that behavioral changes in earlier stages of the task affect set formation thus performance in later stages [58], we administered CNO in both test and control groups just prior to EDS. LC perturbation impaired switching behavior as test group mice took more trials to reach performance criterion (Fig. 3b, trials to reach performance criterion: control vs. test, 14 ± 2 trials vs. 23 ± 3 trials, P = 0.029). Video analysis found no significant difference in overall locomotion activity or reaction time between test and control groups (Supp. Fig. 3), suggesting that the behavioral impairment was not due to apparent changes in motivation or motor functions.

LC inhibition augments single-neuron representation. To assess the neurophysiological effects of LC inhibition on mPFC activity, we simultaneously induced the expression of Gi-DREADD in the LC and GCaMP6f in the mPFC (Fig. 3c). We monitored mPFC activity in mice performing AST while inhibiting LC just prior to EDS (593 neurons from 4 control mice and 446 neurons from 5 test mice). We first evaluated the overall effect of LC inhibition on calcium activity of mPFC neurons. The frequency of calcium transients did not differ between test and control groups, but the test group exhibited a small (~5%) reduction in transient duration (Supp. Fig. 4). Next, we examined how single-neuron encoding was affected by LC inhibition. Somewhat unexpectedly, we found a greater portion of mPFC neurons responding to task-related variables in the test group (Fig. 3d, control vs. test, overall fraction of modulated neurons: 27% (159/593) vs 40% (178/446), P < 1e-3; choice: 10% (59/593) vs. 17% (75/446), P < 1e-3; history: 6% (34/593) vs. 13 % (57/446), P < 1e-3; rule: 17%

(102/593) vs. 25% (111/446), P = 2.4e-3. Chi-squared test). The fraction of mixed-tuning neurons was also augmented with LC inhibition (20% (31/159) vs. 33% (59/178), P = 4.7e-3).

LC inhibition impedes population dynamics and encoding. We moved on to population analysis and hypothesized that impaired ability to rule switching in EDS is associated with disrupted mPFC dynamic changes. Indeed, the same dimensionality-reduction analysis revealed that low-dimensional population vectors representing the two inferred states were less separable when LC was inhibited (Fig. 3e-h, Euclidean distance: control vs. test, 12.8 ± 0.05 vs. 8.9 ± 0.03, P < 1e-3). In addition, the dynamic fluctuations during decision making were reduced (Fig. 3g, i, peak distance, control vs. test: 4.1 ± 0.14 vs. 2.0 ± 0.07, P < 1e-3). This result was consistent with the observation that LC inhibition rendered the population trajectories representing the two states more similar (Fig. 3j, Supp. Fig. 5. Correlation coefficient, control vs. test: 0.15 ± 0.03 vs. 0.95 ± 0.01, P < 1e-3). Together, our results show that LC inhibition dampens and impedes mPFC dynamic processes during rule switching. Based on these findings, it is perhaps not surprising that decoder performance was reduced when LC was inhibited. Specifically, HMM was less accurate in predicting state changes (Fig. 4a, b, state prediction accuracy, control vs. test: 0.72 ± 0.02 vs. 0.55 ± 0.03, P < 1e-3; Δ switch point, control vs. test: -3 ± 1 trials vs. -9 ± 1 trials, P < 1e-3), and GLM was also less accurate in predicting trial-by-trial decisions (Fig. 4c, d, control vs. test: 0.81 ± 0.01 vs. 0.74 ± 0.01, P < 1e-3). Together, our data show that LC inhibition produces a marked impairment in mPFC dynamics and encoding properties, which parallels the behavioral effects.

Direct LC input to the mPFC is involved in rule switching. So far, we showed that LC inhibition impaired switching behavior and impeded mPFC dynamic changes. Given that LC neurons profusely innervate many brain structures [59–61], we wanted to determine whether direct LC input to the mPFC was important for set shifting. To specifically perturb the direct LC-mPFC pathway (Fig. 4e), we expressed Gi-DREADD in the LC but locally infused CNO in the mPFC via bilateral cannula implants just prior to EDS (Fig. 4f). Control group mice also had cannula implants and received CNO injections in the same manner. Consistent with our previous observations from nonspecific LC perturbation (Fig. 3b), inhibiting LC input to the mPFC induced similar behavioral impairment where test group mice required more trials to switch to the new rule (Fig. 4g. Trial to reach the criterion, control vs. test: 16 ± 1 trials vs. 24 ± 2 trials, P = 2.0e-3). Video analysis suggested that the behavioral impairment was not due to apparent motor or motivational changes (Supp. Fig. 6). Supporting previous work [25,54,55], our results provide further evidence for the direct involvement of the LC-mPFC pathway in set shifting.

Discussion

Numerous studies have shown the requirement of the PFC in behavioral flexibility in different model systems including rodents, non-human primates, and human subjects (e.g., [10,19,20,31,62,63]). Our current work is aimed to investigate the neurophysiological substrates underlying PFC's engagement in set shifting. We record mPFC activity in mice learning to discriminate complex stimulus features comprising perceptual dimensions of digging medium and odor. We identify a remarkable dynamic shift in mPFC population representation during rule switching. Pronounced prefrontal encoding capacity is associated with switching across, but not within perceptual dimensions. Inhibiting LC activity affects switching behavior, single-neuron representation, and population dynamics and encoding. The collective findings provide strong evidence to support the unique role of the mPFC in mediating set-shifting processes, and highlight the profound impact of ascending modulatory input from the LC on shaping mPFC dynamics and behavioral transitions.

Loss-of-function work has long pointed to the importance of the mPFC in EDS-type switching [15,19,20,30,31]. However, the neurophysiological mechanisms underlying this specific involvement is not clear. Being the more difficult stages in AST, REV and EDS are designed to address different cognitive processes. In REV, subjects are challenged to flip their responses to the same set of cues (e.g., to promote response to the previously irrelevant cue and to inhibit responses to the previously relevant cue). In EDS, subjects learn to generalize abstract perceptual dimensions, and shift their responses to the change of dimensions instead of to any specific cues [19,20,27–29]. Importantly, the design of EDS is also different from other set-shifting tasks that do not implement the step of

generalizing perceptual dimension-based rules [25,26]. To understand mPFC's specific involvement in AST, we contrast its encoding properties in EDS to REV. To our knowledge, we provide one of the earliest pieces of neurophysiological evidence to shed light on mPFC's involvement in set-shifting behavior. A recent study trained mice to perform a head-fixed set-shifting task and showed that mPFC encodes trial history [34]. These mice were trained for a prolonged period and neuronal representation could shift when behavioral strategies became more routine [64]. In contrast, in the more ethologically relevant, freely moving AST animals had no prior experience of the rule changes and needed to adapt their behavior 'on the fly' [19–22]. It is thus possible that multiple factors, including behavioral context, learning, and task epoch could affect mPFC response and encoding.

Set shifting and flexible behavior involve complex cognitive processes, which likely engage multiple brain areas interacting with the PFC [2,49–53]. Prior studies support the notion that the LC and its input to the PFC play a critical role in attention, working memory and switching [24,25,54–56,65–68]. Our analysis revealed that LC inhibition augmented single-neuron representation in the mPFC. This observation may appear counterintuitive at first glance, but studies have shown that imbalance in the interplay between excitation and inhibition can exert deleterious effects on cognitive functions [69–72]. Specifically, an amplification in neuronal response in one brain region could induce the transmission of noisy information to downstream circuits. Furthermore, the potential link between the augmented mixed tuning at the single-cell level and impaired encoding at the population level is intriguing. It is plausible that an appropriate level of mixed tuning is optimal for population coding, and that too much mixed tuning could deteriorate population representation of task- and decision-related features. Future experimental and computational work will elucidate this potentially important connection.

Norepinephrine (NE) exerts both excitatory and inhibitory influences through distinct types of adrenergic receptors expressed in different cell types [59]. By preferentially binding to specific types of adrenergic receptors in a concentration-dependent way, NE is proposed to mediate downstream neuronal activity and behavior in a non-linear manner [50]. Interestingly, a recent study in the orbital prefrontal cortex showed that the reduction of NE and downregulation of alpha-1 receptors led to a decrease in the activity of GABAergic interneurons, which was important for driving behavioral transitions [68]. Other studies have indicated that LC-NE neurons can co-release other neurotransmitters such as dopamine and peptides [59,73–77]. Together, they highlight intricate mechanisms underlying LC modulation of PFC functions and future studies are needed to better understand the involvement of specific cell types and receptors in flexible decision making.

Our population analysis revealed a pronounced reduction of the dynamic changes in mPFC activity with LC inhibition. This observation suggests attenuated mPFC engagement during the switching processes. Furthermore, the decrement in decoding capacity, coupled with an enhanced mismatch between neural and behavioral switching, underscore LC signaling as an important modulatory factor for mPFC processing and behavior. Disruptions in noradrenergic functions could play a role in the onset or manifestation of stress-, attention- and anxiety-related neuropsychiatric conditions [59,78–81]. These conditions often entail alterations in the functioning of the PFC [57,82]. Supporting previous work showing that LC perturbation leads to changes in PFC response and connectivity [83–87], our data suggest that LC induces notable alterations in mPFC dynamics, affecting its role in orchestrating switching behavior.

Methods

Surgery

All experimental procedures were approved by the UC Riverside Animal Care and Use Committee (AUP20220030). A total of sixty-two mice, aged 8-16 weeks, of mixed sex, were included in the study. Mice were C57BL/6J and DBH-Cre (B6.FVB(Cg)-Tg(Dbh-cre)KH212Gsat/Mmucd, 036778-UCD, MMRRC), singly housed in a vivarium with a reversed light-dark cycle (9a-9p). Twenty-four mice were used in Fig. 1c. Four mice were implanted with GRIN lens to record mPFC activity during AST (Fig.

1d-h, Fig. 2). Four control and six test mice were used for LC manipulation (Fig. 3a, b). Four control and five test mice were implanted with GRIN lens with LC manipulation (Fig. 3c-j, Fig. 4a-d). Seven control and eight test mice were used for LC terminal inhibition (Fig. 4f, g). All surgical procedures were conducted under aseptic conditions, maintaining body temperature with a heating pad. Anesthesia was induced using a mixture of isoflurane (2–3%), and mice were positioned within a precise digital small-animal stereotaxic apparatus (Kopf Instruments and RWD). Before surgery, hair was gently removed from the dorsal head area, ophthalmic ointment was applied to protect the eyes, and the incision site was sanitized with betadine. All measurements were referenced to bregma for virus/implant surgeries. Viral injections were accomplished using a microinjection needle coupled with a 10 µl microsyringe (WPI). The virus was delivered at a controlled rate of 0.03 µl/min via a microsyringe pump (WPI). Following the completion of the injection, a 10-minute interval was allowed before slowly withdrawing the needle. Subsequent to viral infusions, nylon sutures were used to close the incision. Animals received 0.10 mg/kg buprenorphine and 0.22% enofloxacin and were placed in their respective home cages over a heating pad at 37°C. After full recovery from anesthesia, the subjects were returned to vivarium.

For LC inhibition, dopamine-β-hydroxylase (DBH)-Cre mice received bilateral injections of AAV carrying Gi-DREADD receptors (AAV-EF1a-DIO-hM4D(Gi)-mCherry) into the LC (AP: -5.1, ML: 0.95, DV: -3.5 and -3.7 mm, 0.3 µl each depth). For calcium imaging, AAV containing GCaMP6f (AAV-Syn-GCaMP6f-WPRE-SV40) was injected into the medial prefrontal cortex (Prelimbic; AP: 1.8, ML: 0.3, DV: -2.0 and -2.4 mm, 0.3 µl). Following injections, a 30-gauge needle was inserted to create space and reduce tissue resistance to facilitate lens insertion; however, no tissue was aspirated. A gradient refractive index lens (GRIN lens, Inscopix) with a diameter of 0.5 mm and approximately 6 mm in length was gradually lowered through the craniotomy, allowing tissue decompression. This lens was positioned to target AP: -1.8, ML: 0.3, DV: 2.2 mm. The same coordinates were applied for the cannula implantation used in LC terminal inhibition. Lens implants were securely attached to the skull using a layer of adhesive cement (C&B Metabond, Parkell), followed by dental cement (Ortho-Jet, Lang Dental). To protect the lens, a layer of silicone rubber was applied as a protective cover.

Following virus incubation, mice were once again anesthetized under isoflurane and securely positioned in the stereotaxic setup. Baseplates were affixed around the GRIN lens to provide structural support for the attachment of the miniaturized microscope. The top surface of the exposed GRIN lens was meticulously cleaned using a cotton-tipped applicator dipped in a solution of 15% isopropyl alcohol diluted with ddH2O. A miniaturized microscope, equipped with a single-channel epifluorescence and a 475-nm blue LED (Inscopix), was then carefully positioned over the implanted GRIN lens. Adjustments were made along the dorsal-ventral axis to achieve the optimal focal plane for imaging. Subsequently, the microscope/baseplate assembly was secured in place using adhesive cement. The microscope was detached from the baseplates, and a final layer of dental cement was applied to prevent light leakage. A protective plate was placed over the implant until imaging sessions. Mice were singly housed after lens implant.

Behavior and data acquisition

To assess flexible decision-making in freely moving mice, we adopted the 5-stage testing paradigm of the attentional set-shifting task (AST) [21,22]. Two weeks before training, mice were food restricted (85% of initial weight) and handled by the experimenter for 5-7 days. Next, mice were acclimated to the behavioral box (25 x 40 cm) and experimental setup for 3-4 days, followed by a brief training session to stimulate the innate burrowing/digging behavior to retrieve food reward from the ramekins. Two ramekins were placed at two corners of the behavioral box, both containing 25 mg of cheerios. Throughout the training session the reward was gradually buried in clean home cage bedding. In each trial mice were allowed 3-4 minutes to explore. Mice were considered well trained once they could consistently dig and retrieve the reward from both locations for 15-20 trials.

The AST consisted of the following stages: 1) simple discrimination (SD), in which animals choose between two digging mediums associated with distinct textures (first dimension), only one of the two

stimuli predicts food reward; 2) compound discrimination (CD), in which a second stimulus dimension (two odor cues) is explicitly introduced. Each odor cue is randomly paired with a digging medium in every trial, but the reward is still predicted as in SD; 3) intra-dimensional reversal (REV), which preserves the task-relevant dimension (digging medium) but swaps cue contingencies; 4) intra-dimensional shift (IDS), which preserves the task-relevant dimension (digging medium), but replaces all four cues with novel ones (a new digging medium predicts reward); 5) extra-dimensional shift (EDS), which swaps the previous task-relevant and task-irrelevant dimensions with all cues replaced (a new odor cue predicts reward). All stages were performed within a single day, lasting 3-4 hours. In each trial, the ramekin associated with the relevant stimulus contained a retrievable reward. To avoid the possibility that mice used food odor cues to solve the task, the other ramekin contained a non-retrievable reward (trapped under a mesh wire at the bottom). The two ramekins were placed randomly in the two locations every trial. Mice were allowed to complete a trial (dig one ramekin) within 3 minutes. Once mice started digging, the other ramekin was immediately removed from the behavioral box. To reach the criterion the animal has to dig in the correct ramekin six times consecutively and correctly rejecting the incorrect ramekin on at least two occasions.

An overhead CCD camera (Basler acA1300-200um) was set to capture behavior at 20 Hz, controlled by Pylon Viewer Software. Video and calcium recordings were synchronized via a common TTL pulse train (Arduino). Behavioral annotations were done manually post hoc. On the recording day, mice were attached to the miniaturized microscope. Grayscale images were collected at 20 frames per second using 0.01 mW/mm^2 of excitation light. Snout, head, tail, and ear tracking were measured using DeepLabCut [88]. The network was initially trained with 100 uniformly distributed frames from 5 videos, followed by an additional iteration to rectify outlier detections. The measurements for distance and speed were computed using the head, where the likelihood of accuracy exceeded 95 percent. After the test, animals were allowed to access food and water *ad libitum* for 3 days before to be transcardially perfused. Following dissection, brains were post-fixed for 24 h at 4°C in 4% PFA, and sectioned for immunohistochemistry to label TH+ neurons.

Locus coeruleus inactivation

On the test day, Clozapine-N-oxide (CNO) (Sigma) was freshly prepared for systemic or local infusions. For systemic injection we used a concentration of 0.03 mg/Kg diluted in saline. CNO was injected 60 min before EDS test in both test (Gi-DREADD) and control (wild type) group mice. For LC terminal inhibition we used a CNO concentration of 0.5 mM diluted in cortex buffer. Mice were bilaterally implanted with stainless steel cannula guide (26 gauge; RWD) targeting the mPFC. Dummy cannulas were used to fill the guide and removed only during the injection period. CNO was infused at a rate of 0.03 µl/min. After infusion, injecting cannulas were left in place for 5 min to allow drug diffusion.

Image processing

Image processing was executed using Inscopix data processing software (version 1.6). Prior to data analysis, raw imaging videos underwent preprocessing, including a x4 spatial down sampling to reduce file size and processing time. No temporal down sampling was performed. The images were then cropped to eliminate post-registration borders and areas where cells were not visible. Prior to the calculation of the dF/F0 traces, lateral movement was corrected. For ROI identification, we used a constrained non-negative matrix factorization algorithm optimized for endoscopic data (CNMF-E) to extract fluorescence traces from ROIs. The detected ROIs were then manually evaluated based on neuronal morphology, ensuring adequate separation from neighboring cells.

Single unit analysis

Calcium signals for each ROI were z-scored and aligned to behavioral events (i.e., trial start, digging) using MATLAB (MathWorks). In order to classify neuronal representations of different task-related variables, we performed Receiver-Operating-Characteristic (ROC) analysis [89] on the activity of each

neuron prior to decision making. Task rule representation was defined as significant calcium responses between early (trial-and-error learning) and late (rule-following) trials during a time window prior to decision (-5 to 0 s). Similar analysis was performed to classify trial history encoding, comparing calcium activity during the same time window (-5 to 0 s to from decision) after correct trials against after incorrect trials; and choice encoding, comparing calcium activity (-5 to 0 s from decision) when the choice in the current trial is correct or incorrect.

Dimensionality reduction

We concatenated neuronal activity across recordings, and constructed population vector (PV(t)) in the early and late states by assembling individual calcium values (in 50-ms bins, without overlap) over a period of ten seconds centered at the decision point (digging). These calcium values were extracted from 80% of recorded neurons randomly selected. This process was reiterated 20 times to account for the inherent variability in the dynamics of the population vector. The resultant high-dimensional trajectories were smoothed and embedded into a lower-dimensional space through principal component analysis (PCA). The explained variance was calculated for the first six principal components, which collectively accounted for over 80% of the total variance. Projections into a low-dimensional space (n = 3) were generated for visualization purposes. To explore potential variations in the modulation of the neuronal population between the two behavioral states, we computed the Euclidean distance between these two vectors. As a control, we conducted 20 iterations of trial sequence shuffling to construct the population vector. Additionally, to evaluate the degree of similarity between vectors in each iteration, Pearson's correlation was computed for the time series of individual principal components. Subsequently, these correlation coefficients were averaged to derive an overall measure of vector similarity. Subsequently, we performed Euclidean distance and Pearson's correlation analyses.

Hidden Markov model

To predict behavioral state transitions using neural data, we employed a hidden Markov model (HMM) approach [38]. This method enabled us to capture these observable shifts by analyzing neural population activity on a trial-by-trial basis for each animal. In this context, the behavioral sequence was represented as a series of discrete states, with each state corresponding to a specific population neuronal activity (early - 0 and late - 1). Neuronal states were mapped as observations or emissions associated with each behavioral state. Essentially, whenever the animal entered a particular state, it was linked to a corresponding distribution of neuronal states associated with that particular state. To construct the distribution of neuronal states, we bootstrapped 200 times, each time 80% of recorded neurons, averaging their activity in the five seconds period preceding the animal's decision in each trial. Subsequently, a principal component analysis was conducted to generate a high-dimensional neuronal distribution of the activity. Employing a k-means algorithm, we clustered the neuronal population, focusing on the initial three principal components. Our objective was to identify potential associations with distinct types of activity, specifically categorized as 'engage-win,' 'engage-lose,' and 'no-engage.' Subsequently, we computed the emission and transition probabilities during the early and late states. The HMM was trained on 90% of the bootstrapped dataset, and the model's performance was tested on the remaining 10% using the Viterbi algorithm. Model accuracy was determined by comparing the actual behavioral sequence with the sequence generated by the model on the testing data. Furthermore, the switch point (the location of the first trial in the late state) was determined by calculating the difference between the actual behavioral trial number and the model's trial number. As a control, we implemented a trial shuffling of neuronal sequence distributions for training the HMM and a noticeable decline in performance was observed (Model accuracy data vs. control: P = 7.8e-5, data not shown). This control was essential to establish a baseline for model performance in the absence of meaningful data patterns.

Generalized linear model

We conducted a multivariate regression analysis on the population vectors to elucidate the role of the mPFC in the decision-making processes. The regression model aimed to predict trial outcomes (correct or incorrect) based on population activity patterns. To construct the population vector for each session, we initially computed the average activity of the recorded neurons in the 5 seconds prior to decision. We then randomly selected 80% of the recorded neurons and a principal component analysis was applied to the population vector. The first three principal components were retained as predictors for the regression model. To address variability and ensure robustness, we conducted one hundred bootstrap procedures. Subsequently, we partitioned 80% of the dataset for model training and tested the model on the remaining 20% of unseen data. We established a threshold of 0.5 to binarize the model probability. Values above 0.5 were assigned a label of 1, while values below 0.5 were assigned a label of 0. Model accuracy was assessed by comparing the actual behavioral sequence with the sequence generated by the model. As a control measure, we randomized the label of each trial and trained and tested the model and a significant decrease in the model's performance was observed ($P = 1.3e-6$).

Statistical analysis

Data were reported as mean ± SEM unless otherwise noted. We did not use statistical methods to predetermine sample sizes. Sample sizes were similar to those reported in the field. We assigned mice to experimental groups arbitrarily, without randomization or blinding. Unless otherwise noted, statistical tests were two-tailed Wilcoxon rank-sum when sample sizes were >7. When sample sizes were ≤ 7, two-tailed t tests were used.

Author contributions

M.N., L.S.T. and H.Y. planned the project and built the apparatus. M.N. performed experiments. L.S.T. assisted with LC inhibition experiments, M.G. assisted with histology, K.D. assisted with behavioral video analysis, and N.E.Z. assisted with data analysis. M.N. and H.Y. analyzed data and wrote the manuscript with input from L.S.T., M.G. and N.E.Z.

Acknowledgements

We thank Shaorong Ma for helping with the behavioral paradigm, and Edward Zagha for commenting on the manuscript. H.Y. was supported by Klingenstein-Simons Fellowship Awards in Neuroscience, and NIH grants R01NS107355 and R01NS112200.

References:

1. Miller, E. K. & Cohen, J. D. An Integrative Theory of Prefrontal Cortex Function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
2. Uddin, L. Q. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat. Rev. Neurosci.* (2021) doi:10.1038/s41583-021-00428-w.
3. Mesulam, M. M. From sensation to cognition. *Brain* **121**, 1013–1052 (1998).
4. Miller, E. K. The prefrontal cortex: Complex neural properties for complex behavior. *Neuron* **22**, 15–17 (1999).
5. Ragozzino, M. E. The contribution of the medial prefrontal cortex, orbitofrontal cortex, and dorsomedial striatum to behavioral flexibility. *Ann. N. Y. Acad. Sci.* **1121**, 355–375 (2007).
6. Le Merre, P., Ährlund-Richter, S. & Carlén, M. The mouse prefrontal cortex: Unity in diversity. *Neuron* 1–20 (2021) doi:10.1016/j.neuron.2021.03.035.
7. Berg, E. A. A simple objective technique for measuring flexibility in thinking. *J. Gen. Psychol.* **39**, 15–22 (1948).
8. Milner, B. Effects of Different Brain Lesions on Card Sorting. *Arch. Neurol.* **9**, 90 (1963).
9. Roberts, A. C., Robbins, T. W. & Everitt, B. J. The Effects of Intradimensional and Extradimensional Shifts on Visual Discrimination Learning in Humans and Non-human Primates. *Q. J. Exp. Psychol. Sect. B* **40**, 321–341 (1988).
10. Dias, R., Robbins, T. W. & Roberts, A. C. Primate analogue of the Wisconsin card sorting test: Effects of excitotoxic lesions of the prefrontal cortex in the marmoset. *Behav. Neurosci.* **110**, 872–886 (1996).
11. Monchi, O., Petrides, M., Petre, V., Worsley, K. & Dagher, A. Wisconsin card sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J. Neurosci.* **21**, 7733–7741 (2001).
12. Barnett, J. H. *et al.* Assessing cognitive function in clinical trials of schizophrenia. *Neurosci. Biobehav. Rev.* **34**, 1161–1177 (2010).
13. Brown, V. J. & Tait, D. S. Attentional Set-Shifting Across Species. in 363–395 (2015). doi:10.1007/7854_2015_5002.
14. Young, J. W., Powell, S. B., Geyer, M. A., Jeste, D. V. & Risbrough, V. B. The mouse attentional-set-shifting task: A method for assaying successful cognitive aging? *Cogn. Affect. Behav. Neurosci.* **10**, 243–251 (2010).
15. Dias, R., Robbins, T. W. & Roberts, A. C. Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* **380**, 69–72 (1996).
16. Ridderinkhof, K. R. The Role of the Medial Frontal Cortex in Cognitive Control. *Science (80-.).* **306**, 443–447 (2004).
17. Floresco, S. B., Zhang, Y. & Enomoto, T. Neural circuits subserving behavioral flexibility and their relevance to schizophrenia. *Behav. Brain Res.* **204**, 396–409 (2009).
18. Dajani, D. R. *et al.* Measuring cognitive flexibility with the flexible item selection task: From Fmri adaptation to individual Connectome mapping. *J. Cogn. Neurosci.* **32**, 1026–1045 (2020).
19. Birrell, J. M. & Brown, V. J. Medial frontal cortex mediates perceptual attentional set shifting in the rat. *J. Neurosci.* **20**, 4320–4324 (2000).
20. Bissonette, G. B. *et al.* Double dissociation of the effects of medial and orbital prefrontal cortical lesions on attentional and affective shifts in mice. *J. Neurosci.* **28**, 11124–11130 (2008).
21. Liston, C. *et al.* Stress-Induced Alterations in Prefrontal Cortical Dendritic Morphology Predict Selective Impairments in Perceptual Attentional Set-Shifting. *J. Neurosci.* **26**, 7870–7874 (2006).
22. Snyder, K., Wang, W. W., Han, R., McFadden, K. & Valentino, R. J. Corticotropin-releasing factor in the norepinephrine nucleus, locus coeruleus, facilitates behavioral flexibility. *Neuropsychopharmacology* **37**, 520–530 (2012).
23. Bissonette, G. B., Powell, E. M. & Roesch, M. R. Neural structures underlying set-shifting: Roles of medial prefrontal cortex and anterior cingulate cortex. *Behav. Brain Res.* **250**, 91–101 (2013).
24. Lapiz, M. D. S., Bondi, C. O. & Morilak, D. A. Chronic treatment with desipramine improves

cognitive performance of rats in an attentional set-shifting test. *Neuropsychopharmacology* **32**, 1000–1010 (2007).
25. Cope, Z. A., Vazey, E. M., Floresco, S. B. & Aston Jones, G. S. DREADD-mediated modulation of locus coeruleus inputs to mPFC improves strategy set-shifting. *Neurobiol. Learn. Mem.* **161**, 1–11 (2019).
26. Cho, K. K. A. *et al.* Cross-hemispheric gamma synchrony between prefrontal parvalbumin interneurons supports behavioral adaptation during rule shift learning. *Nat. Neurosci.* **23**, 892–902 (2020).
27. Garner, J. P., Thogerson, C. M., Würbel, H., Murray, J. D. & Mench, J. A. Animal neuropsychology: Validation of the Intra-Dimensional Extra-Dimensional set shifting task for mice. *Behav. Brain Res.* **173**, 53–61 (2006).
28. Bissonette, G. B. & Powell, E. M. Reversal learning and attentional set-shifting in mice. *Neuropharmacology* **62**, 1168–1174 (2012).
29. Colacicco, G., Welzl, H., Lipp, H. P. & Würbel, H. Attentional set-shifting in mice: Modification of a rat paradigm, and evidence for strain-dependent variation. *Behav. Brain Res.* **132**, 95–102 (2002).
30. McAlonan, K. & Brown, V. J. Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behav. Brain Res.* **146**, 97–103 (2003).
31. Dias, R., Robbins, T. W. & Roberts, A. C. Dissociable Forms of Inhibitory Control within Prefrontal Cortex with an Analog of the Wisconsin Card Sort Test: Restriction to Novel Situations and Independence from “On-Line” Processing. *J. Neurosci.* **17**, 9285–9297 (1997).
32. Mansouri, F. A., Matsumoto, K. & Tanaka, K. Prefrontal cell activities related to monkeys’ success and failure in adapting to rule changes in a Wisconsin card sorting test analog. *J. Neurosci.* **26**, 2745–2756 (2006).
33. Norman, K. J. *et al.* Post-error recruitment of frontal sensory cortical projections promotes attention in mice. *Neuron* **109**, 1202-1213.e5 (2021).
34. Spellman, T., Svei, M., Kaminsky, J., Manzano-Nieves, G. & Liston, C. Prefrontal deep projection neurons enable cognitive flexibility via persistent feedback monitoring. *Cell* **184**, 2750-2766.e17 (2021).
35. Lui, J. H. *et al.* Differential encoding in prefrontal cortex projection neuron classes across cognitive tasks. *Cell* **184**, 489-506.e26 (2021).
36. Jercog, D. *et al.* Dynamical prefrontal population coding during defensive behaviours. *Nature* (2021) doi:10.1038/s41586-021-03726-6.
37. Del Arco, A., Park, J., Wood, J., Kim, Y. & Moghaddam, B. Adaptive encoding of outcome prediction by prefrontal cortex ensembles supports behavioral flexibility. *J. Neurosci.* **37**, 8363–8373 (2017).
38. Durstewitz, D., Vittoz, N. M., Floresco, S. B. & Seamans, J. K. Abrupt Transitions between Prefrontal Neural Ensemble States Accompany Behavioral Transitions during Rule Learning. *Neuron* **66**, 438–448 (2010).
39. Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
40. Kim, H., Ährlund-Richter, S., Wang, X., Deisseroth, K. & Carlén, M. Prefrontal Parvalbumin Neurons in Control of Attention. *Cell* **164**, 208–218 (2016).
41. Pinto, L. & Dan, Y. Cell-Type-Specific Activity in Prefrontal Cortex during Goal-Directed Behavior. *Neuron* **87**, 437–450 (2015).
42. Yuste, R. From the neuron doctrine to neural networks. *Nat. Rev. Neurosci.* **16**, 487–497 (2015).
43. Pouget, a, Dayan, P. & Zemel, R. Information processing with population codes. *Nat. Rev. Neurosci.* **1**, 125–32 (2000).
44. Ebitz, R. B. & Hayden, B. Y. The population doctrine in cognitive neuroscience. *Neuron* 1–14 (2021) doi:10.1016/j.neuron.2021.07.011.
45. Saxena, S. & Cunningham, J. P. Towards the neural population doctrine. *Curr. Opin. Neurobiol.* **55**, 103–111 (2019).
46. Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K. & Poggio, T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* **100**,

1407–1419 (2008).
47. Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N. & Harvey, C. D. Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. *Cell* **170**, 986-999.e16 (2017).
48. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: High dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **37**, 66–74 (2016).
49. Wimmer, R. D. *et al.* Thalamic control of sensory selection in divided attention. *Nature* **526**, 705–709 (2015).
50. Aston-Jones, G. & Cohen, J. D. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28**, 403–50 (2005).
51. Kaefer, K., Nardin, M., Blahna, K. & Csicsvari, J. Replay of Behavioral Sequences in the Medial Prefrontal Cortex during Rule Switching. *Neuron* **106**, 154-165.e6 (2020).
52. Sadacca, B. F., Wikenheiser, A. M. & Schoenbaum, G. Toward a theoretical role for tonic norepinephrine in the orbitofrontal cortex in facilitating flexible learning. *Neuroscience* **345**, 124–129 (2017).
53. Cerpa, J. C., Coutureau, E. & Parkes, S. L. Dopamine and noradrenaline modulation of goal-directed behavior in orbital and medial prefrontal cortex: Toward a division of labor? *Behav. Neurosci.* **135**, 138–153 (2021).
54. Lapiz, M. D. S. & Morilak, D. A. Noradrenergic modulation of cognitive function in rat medial prefrontal cortex as measured by attentional set shifting capability. *Neuroscience* **137**, 1039–1049 (2006).
55. McGaughy, J., Ross, R. S. & Eichenbaum, H. Noradrenergic, but not cholinergic, deafferentation of prefrontal cortex impairs attentional set-shifting. *Neuroscience* **153**, 63–71 (2008).
56. Janitzky, K. *et al.* Optogenetic silencing of locus coeruleus activity in mice impairs cognitive flexibility in an attentional set-shifting task. *Front. Behav. Neurosci.* **9**, 1–8 (2015).
57. Arnsten, A. F. T., Wang, M. J. & Paspalas, C. D. Neuromodulation of Thought: Flexibilities and Vulnerabilities in Prefrontal Cortical Network Synapses. *Neuron* **76**, 223–239 (2012).
58. Chase, E. A., Tait, D. S. & Brown, V. J. Lesions of the orbital prefrontal cortex impair the formation of attentional set in rats. *Eur. J. Neurosci.* **36**, 2368–2375 (2012).
59. Berridge, C. W. & Waterhouse, B. D. The locus coeruleus-noradreneergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Res. Rev.* **42**, 33–84 (2003).
60. Schwarz, L. A. & Luo, L. Organization of the locus coeruleus-norepinephrine system. *Curr. Biol.* **25**, R1051–R1056 (2015).
61. Schwarz, L. A. *et al.* Viral-genetic tracing of the input-output organization of a central noradrenaline circuit. *Nature* **524**, 88–92 (2015).
62. Hampshire, A. & Owen, A. M. Fractionating attentional control using event-related fMRI. *Cereb. Cortex* **16**, 1679–1689 (2006).
63. Owen, A. M., Roberts, A. C., Hodges, J. R. & Robbins, T. W. Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease. *Brain* **116**, 1159–1175 (1993).
64. Chiang, F.-K., Wallis, J. D. & Rich, E. L. Cognitive strategies shift information from single neurons to populations in prefrontal cortex. *Neuron* 1–13 (2021) doi:10.1016/j.neuron.2021.11.021.
65. Devauges, V. & Sara, S. J. Activation of the noradrenergic system facilitates an attentional shift in the rat. *Behav. Brain Res.* **39**, 19–28 (1990).
66. Tait, D. S. *et al.* Lesions of the dorsal noradrenergic bundle impair attentional set-shifting in the rat. *Eur. J. Neurosci.* **25**, 3719–3724 (2007).
67. Newman, L. A., Darling, J. & McGaughy, J. Atomoxetine reverses attentional deficits produced by noradrenergic deafferentation of medial prefrontal cortex. *Psychopharmacology (Berl).* **200**, 39–50 (2008).
68. Li, C. *et al.* A neural circuit for regulating a behavioral switch in response to prolonged uncontrollability in mice. *Neuron* 1–15 (2023) doi:10.1016/j.neuron.2023.05.023.
69. Lam, N. H. *et al.* Effects of Altered Excitation-Inhibition Balance on Decision Making in a

Cortical Circuit Model. *J. Neurosci.* **42**, 1035–1053 (2022).
70. Selimbeyoglu, A. *et al.* Modulation of prefrontal cortex excitation/inhibition balance rescues social behavior in CNTNAP2-deficient mice. *Sci. Transl. Med.* **9**, (2017).
71. Wang, Y. *et al.* Noradrenergic lesion of the locus coeruleus increases the firing activity of the medial prefrontal cortex pyramidal neurons and the role of α2-adrenoceptors in normal and medial forebrain bundle lesioned rats. *Brain Res.* **1324**, 64–74 (2010).
72. Aston-Jones, G. & Cohen, J. D. Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *J. Comp. Neurol.* **493**, 99–110 (2005).
73. Holets, V. ., Hökfelt, T., Rökaeus, Å., Terenius, L. & Goldstein, M. Locus coeruleus neurons in the rat containing neuropeptide Y, tyrosine hydroxylase or galanin and their efferent projections to the spinal cord, cerebral cortex and hypothalamus. *Neuroscience* **24**, 893–906 (1988).
74. Devoto, P. & Flore, G. On the origin of cortical dopamine: is it a co-transmitter in noradrenergic neurons? *Curr. Neuropharmacol.* **4**, 115–25 (2006).
75. Devoto, P., Flore, G., Saba, P., Fà, M. & Gessa, G. L. Stimulation of the locus coeruleus elicits noradrenaline and dopamine release in the medial prefrontal and parietal cortex. *J. Neurochem.* **92**, 368–374 (2005).
76. Devoto, P., Flore, G., Saba, P., Fà, M. & Gessa, G. L. Co-release of noradrenaline and dopamine in the cerebral cortex elicited by single train and repeated train stimulation of the locus coeruleus. *BMC Neurosci.* **6**, 1–11 (2005).
77. Kempadoo, K. A., Mosharov, E. V., Choi, S. J., Sulzer, D. & Kandel, E. R. Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proc. Natl. Acad. Sci.* **113**, 14835–14840 (2016).
78. Briley, M. & Chantal, M. The importance of norepinephrine in depression. *Neuropsychiatr. Dis. Treat.* 9 (2011) doi:10.2147/NDT.S19619.
79. Ressler, K. J. & Nemeroff, C. B. Role of norepinephrine in the pathophysiology and treatment of mood disorders. *Biol. Psychiatry* **46**, 1219–1233 (1999).
80. Herrmann, N., Lanctôt, K. L. & Khan, L. R. The role of norepinephrine in the behavioral and psychological symptoms of dementia. *J. Neuropsychiatry Clin. Neurosci.* **16**, 261–276 (2004).
81. Southwick, S. M. *et al.* Role of norepinephrine in the pathophysiology and treatment of posttraumatic stress disorder. *Biol. Psychiatry* **46**, 1192–1204 (1999).
82. Arnsten, A. F. T. Through the Looking Glass: Differential Noradenergic Modulation of Prefrontal Cortical Function. *Neural Plast.* **7**, 133–146 (2000).
83. Neves, R. M., van Keulen, S., Yang, M., Logothetis, N. K. & Eschenko, O. Locus coeruleus phasic discharge is essential for stimulus-induced gamma oscillations in the prefrontal cortex. *J. Neurophysiol.* **119**, 904–920 (2018).
84. Oyarzabal, E. A. *et al.* Chemogenetic stimulation of tonic locus coeruleus activity strengthens the default mode network. *Sci. Adv.* **8**, 1–13 (2022).
85. Bouret, S. & Sara, S. J. Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. *Eur. J. Neurosci.* **20**, 791–802 (2004).
86. Giustino, T. F., Fitzgerald, P. J., Ressler, R. L. & Maren, S. Locus coeruleus toggles reciprocal prefrontal firing to reinstate fear. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8570–8575 (2019).
87. Zerbi, V. *et al.* Rapid Reconfiguration of the Functional Connectome after Chemogenetic Locus Coeruleus Activation. *Neuron* **103**, 702-718.e5 (2019).
88. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
89. Green, D. M. & Swets, J. A. *Signal detection theory and psychophysics*. (John Wiley and Sons Inc, 1966).

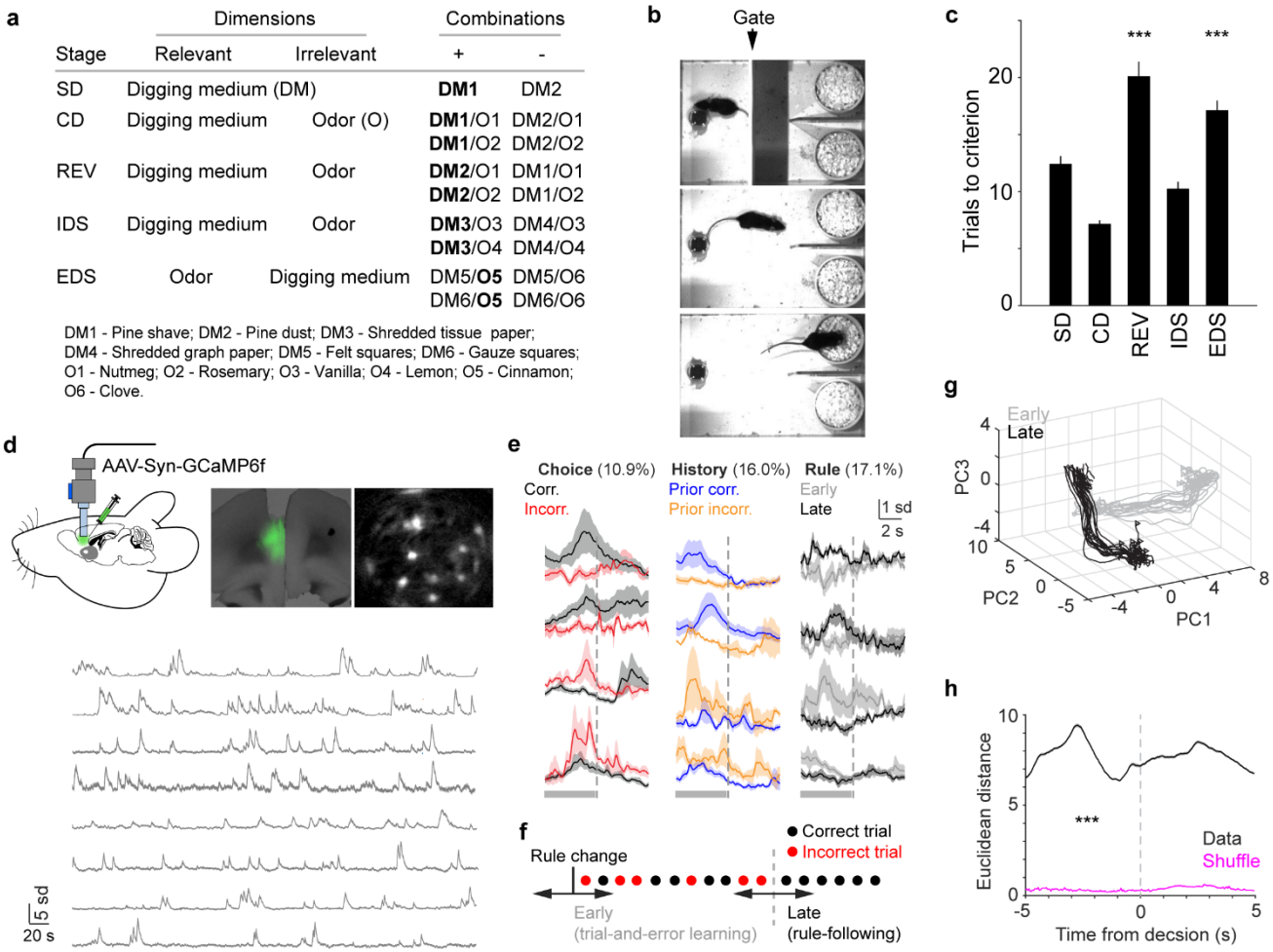

Figure 1. Recording mPFC activity during AST.

(a) Task overview. (b) Example frames (top to bottom) showing when the mouse was in the waiting area, approaching the bowls, and digging. (c) Task performance (total number of trials to criterion) varied across stages (n = 24): SD, 12.4 ± 0.7 trials; CD, 7.2 ± 0.3 trials; REV, 20.1 ± 1.3 trials; IDS, 10.3 ± 0.6 trials; EDS, 17.1 ± 0.9 trials. More trials were required to complete REV and EDS. REV vs. SD, P = 1.0e-4; REV vs. CD: P = 1.5e-5; REV vs. IDS: P = 2.6e-5; EDS vs. SD: P = 1.0e-3; EDS vs. CD: P = 1.8e-5; EDS vs. IDS: P = 2.6e-5. Wilcoxon signed-rank test with Bonferroni correction. (d) Top (left to right): Illustration of miniscope recording in the mPFC; coronal brain section showing lens position and GCaMP6f expression in the prelimbic region; snapshot of miniscope recording during behavior. Bottom: Example time series of fluorescence signals from 8 ROIs. (e) Left to right: Example traces of individual mPFC neurons encoding choice (left), trial history (middle) and task rule (right) based on activity prior to decision (gray bars). (f) Example behavioral progression in the EDS stage. We define the last set of consecutive correct trials (rule-following) and prior mixed correct and incorrect trials (rule-learning) as two different states. (g) PCA analysis revealing high-dimensional trajectories of population activity in the low-dimensional space. Gray and black trajectories represent mPFC activity in the early and late states. Each line represents a population vector from a subset of neurons. (h) Euclidean distance (mean ± SEM) between trajectories in the early and late states aligned to decision. Trials of shuffled labels were used as controls (pink). Euclidean distance quantified in the 5-s window prior to decision, data vs. shuffle: 7.5 ± 0.05 vs. 0.3 ± 0.02, P = 2.6e-34, rank sum = 5050, n = 20. Sample size here represents number of bootstraps.

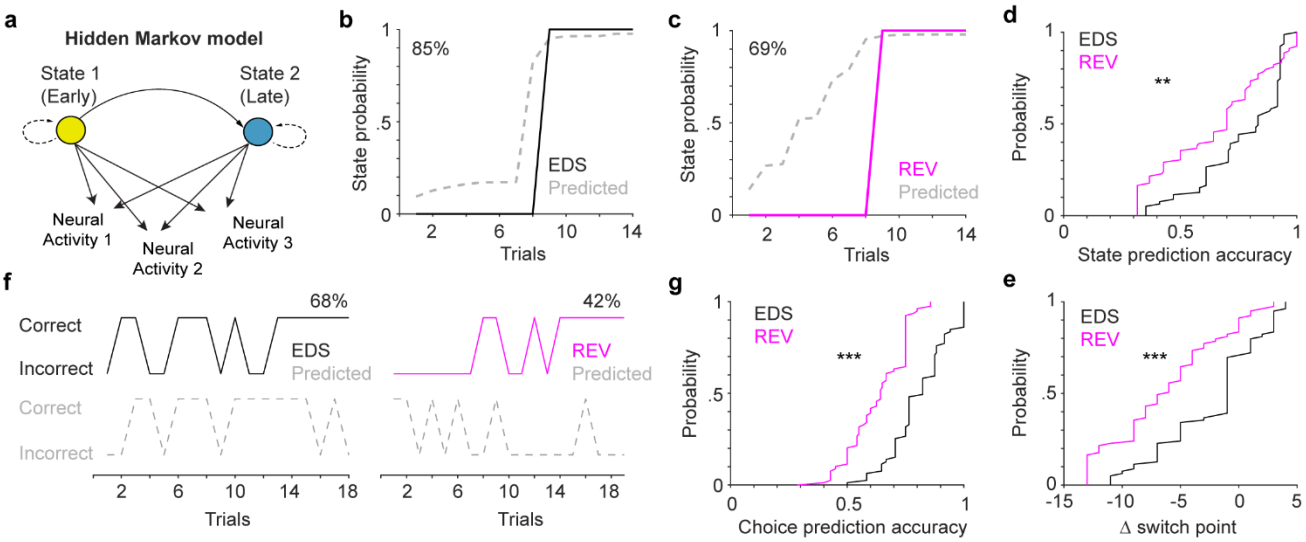

Figure 2. Using mPFC population activity to decode animal behavior.
(a) Schematic of the hidden Markov model to predict the hidden underlying states. (b) An example behavioral state progression in EDS (solid curve: 0-early, 1-late) and HMM predicted state changes (dashed curve). Model prediction accuracy is 85%. (c) An example behavioral state progression in REV (solid) and HMM predicted transition (dashed). Model prediction accuracy is 69%. (d) Cumulative distribution of the accuracy of predicting behavioral states in EDS (black, n = 80) and REV (pink, n = 80). Sample size represents the total number of iterations that the model was tested (20 times per recording, 4 mice). EDS vs. REV: 0.77 ± 0.02 vs. 0.65 ± 0.03, P = 2.0e-3, rank sum = 524.5. (e) Cumulative distribution of the accuracy of predicting state switch point using HMM for EDS (black, n = 80) and REV (pink, n = 80). Δ switch point, EDS vs. REV: -2 ± 1 trials vs. -6 ± 1 trials, P = 8.9e-7, rank sum = 7.9e3. (f) Example sequences of animals' choices (top) and GLM predicted choices (bottom) for an EDS (left) and a REV (right) session. Prediction accuracy is 68% (EDS) and 42% (REV). (g) Cumulative distribution of the accuracy of predicting trial-by-trial choices in EDS (black) and REV (pink). EDS vs. REV: 0.72 ± 0.01 vs. 0.60 ± 0.02, P = 8.3e-5, rank sum = 7.1e3, n = 80.

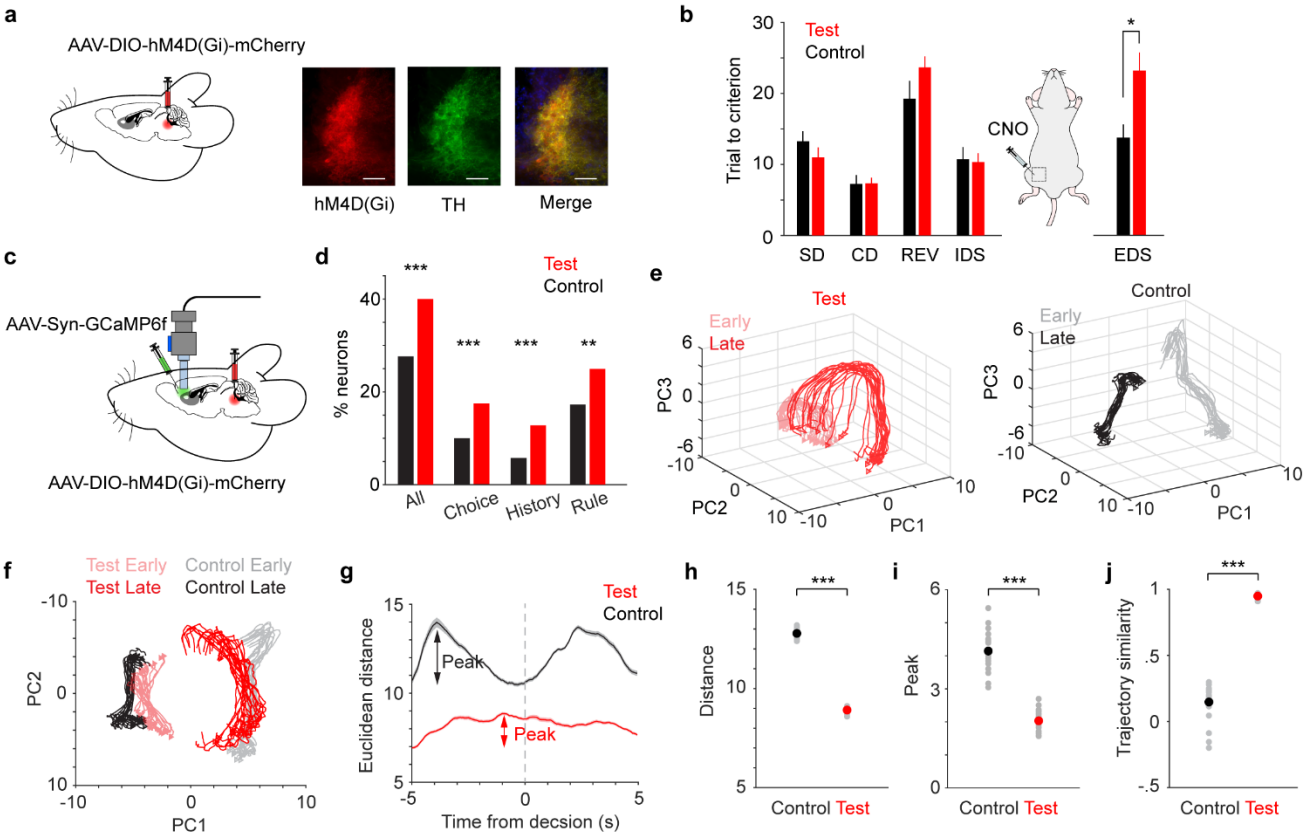

Figure 3. LC inhibition affects behavior and mPFC dynamics

(a) Schematic of DREADD inhibition in the LC and histological images showing DREADD(Gi) and TH expression in the LC of a DBH-Cre mouse. Scalebar: 100 µm. (b) Task performance in the control (n = 4, WT) and test (n = 6) groups. Following CNO injection before EDS, test group mice took more trials to complete this stage (Trial to reach the criterion: control vs. test, 14 ± 2 trials vs. 23 ± 3 trials, P = 0.028, t = -2.7). (c) Schematic of mPFC recording with LC manipulation. (d) Bar plots showing the percentage of mPFC neurons encoding task-related variables in the control (black) and test (red) groups. Control vs. test, overall fraction of modulated neurons: 27% (159/593) vs. 40% (178/446), P = 8.1e-6; Choice: 10% (59/593) vs. 17% (76/446), P = 7.7e-4; History, 6% (34/593) vs. 13%, (57/446), P = 7.0e-5; Rule: 17% (102/593) vs. 25%, (111/446), P = 2.4e-3, Chi-squared test. (e) Population vectors of mPFC activity representing early (light color) and late (dark color) states in test (red) and control (black) groups. (f) Projection of population vectors in (e) onto the first two PCs. (g) Euclidean distance (mean ± SEM) between trajectories representing the two states aligned to decision. Arrows indicate maximal fluctuations prior to decision (peak). (h) Comparison of Euclidean distance quantified prior to decision for test (red) and control (black) groups. Control vs. test, 12.8 ± 0.05 vs. 8.9 ± 0.03, P = 6.8e-8, rank sum = 610, n = 20. Sample size here represents number of bootstraps. (i) Comparison of peak Euclidean distance quantified prior to decision for test (red) and control (black) groups. Control vs. test: 4.1 ± 0.14 vs. 2.0 ± 0.07, P = 6.8e-8, rank sum = 610, n = 20. (j) Comparison of trajectory similarities in the two states. Correlation coefficient, control vs. test: 0.15 ± 0.03 vs. 0.95 ± 0.01, P = 6.8e-8, rank sum = 210, n = 20). Black and red dots indicate group mean in (h-j).

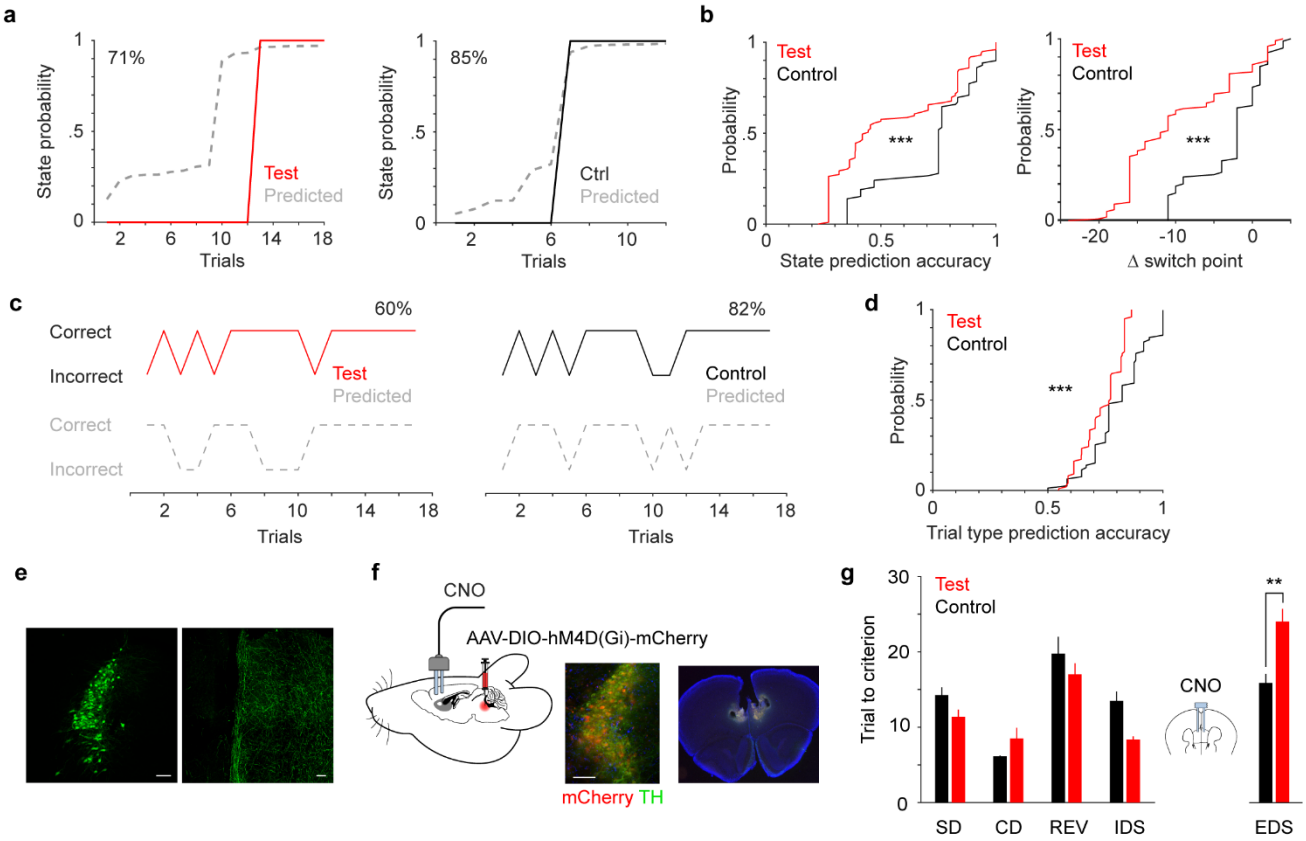

Figure 4. LC inhibition affects mPFC decoding capacity.
(a) Example of behavioral state progressions in an EDS session (solid curve: 0-early, 1-late) and HMM predicted state changes (dashed curve) in test (left) and control (right) conditions. State prediction accuracy is 71% (test) and 85% (control). (b) Left: Cumulative distribution of the accuracy of predicting behavioral states in control (black, n = 80) and test (pink, n = 100) groups. Sample size here represents the total number of iterations that the model was tested (20 times per recording, 4 control mice and 5 test mice). Control vs. test: 0.72 ± 0.02 vs. 0.55 ± 0.03, P = 4.3e-5, rank sum = 8.7e3. Right: Cumulative distribution of the accuracy of predicting switch point in control (black) and test (pink) groups. Control vs. test: -3 ± 1 trials vs. -9 ± 1 trials, P = 5.1e-10, rank sum = 9.4e3. (c) Example sequences of animals' choices (top) and GLM predicted choices (bottom) for a test (left) and a control (right) session. Prediction accuracy is 60% (test) and 82% (control). (d) Cumulative distribution of the accuracy of predicting trial-by-trial choices in control (black, n = 80) and test (red, n = 100) groups. Control vs. test: 0.81 ± 0.01 vs. 0.74 ± 0.01, P = 3.1e-4, rank sum = 8.5e3. (e) Histology showing the expression of EYFP in LC cell bodies (left) and their terminals (right) in the mPFC. AAV-DIO-ChR2-EYFP was injected in the LC of a DBH-Cre mouse. Scalebars: 100 µm. (f) Schematic of inhibiting LC terminals in mPFC and histology displaying DREADD(Gi) expression in the LC and cannula placement in mPFC. (g) Task performance in the control (n = 7, WT) and test (n = 8) groups. Following localized CNO injection before EDS, test group mice took more trials to complete this stage (Trial to reach the criterion, control vs test: 16 ± 1 trials vs. 24 ± 2 trials, P = 2.0e-3, t = -3.9).