

# Distinct roles of monkey OFC-subcortical pathways in adaptive behavior

Kei Oyama<sup>1,6</sup>, Kei Majima<sup>2,6</sup>, Yuji Nagai<sup>1</sup>, Yukiko Hori<sup>1</sup>, Toshiyuki Hirabayashi<sup>1</sup>, Mark A G Eldridge<sup>3</sup>, Koki Mimura<sup>1,4</sup>, Naohisa Miyakawa<sup>1</sup>, Atsushi Fujimoto<sup>1</sup>, Yuki Hori<sup>1</sup>, Haruhiko Iwaoki<sup>1</sup>, Ken-ichi Inoue<sup>5</sup>, Richard C Saunders<sup>3</sup>, Masahiko Takada<sup>5</sup>, Noriaki Yahata<sup>2</sup>, Makoto Higuchi<sup>1</sup>, Barry J Richmond<sup>3</sup>, and Takafumi Minamimoto<sup>1</sup>✉

<sup>1</sup>Department of Functional Brain Imaging, National Institutes for Quantum Science and Technology, Chiba, Japan

<sup>2</sup>Institute for Quantum Life Science, National Institutes for Quantum Science and Technology, Chiba, Japan

<sup>3</sup>Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, USA

<sup>4</sup>Research Center for Medical and Health Data Science, The Institute of Statistical Mathematics, Tachikawa, Japan

<sup>5</sup>Systems Neuroscience Section, Center for the Evolutionary Origins of Human Behavior, Kyoto University, Inuyama, Japan

<sup>6</sup>PRESTO, Japan Science and Technology Agency, Kawaguchi, Japan

**To be the most successful, primates must adapt to changing environments and optimize their behavior by making the most beneficial choices. At the core of adaptive behavior is the orbitofrontal cortex (OFC) of the brain, which updates choice value through direct experience or knowledge-based inference. Here, we identify distinct neural circuitry underlying these two separate abilities. We designed two behavioral tasks in which macaque monkeys updated the values of certain items, either by directly experiencing changes in stimulus-reward associations, or by inferring the value of unexperienced items based on the task's rules. Chemogenetic silencing of bilateral OFC combined with mathematical model-fitting analysis revealed that monkey OFC is involved in updating item value based on both experience and inference. In vivo imaging of chemogenetic receptors by positron emission tomography allowed us to map projections from the OFC to the rostromedial caudate nucleus (rmCD) and the medial part of the mediodorsal thalamus (MDm). Chemogenetic silencing of the OFC-rmCD pathway impaired experience-based value updating, while silencing the OFC-MDm pathway impaired inference-based value updating. Our results thus demonstrate a dissociable contribution of distinct OFC projections to different behavioral strategies, and provide new insights into the neural basis of value-based adaptive decision-making in primates.**

Monkey, Chemogenetics| Orbitofrontal cortex| Caudate nucleus| Mediodorsal thalamus| Pathway-selective silencing| Model-fitting analysis

Correspondence: [minamimoto.takafumi@qst.go.jp](mailto:minamimoto.takafumi@qst.go.jp) (T.M.)

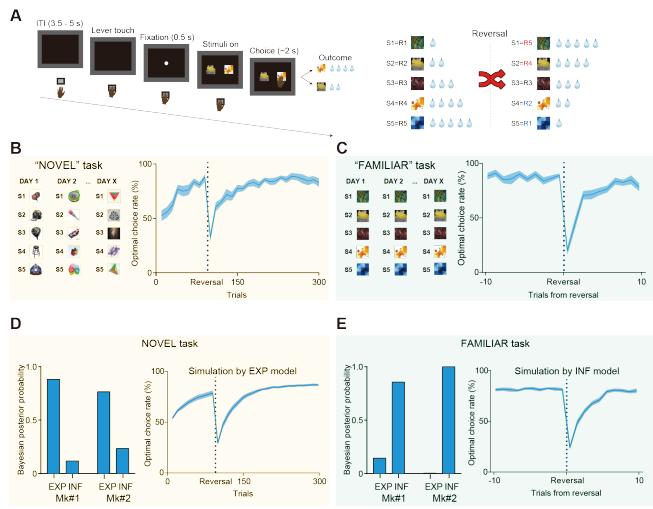
## Introduction

To survive in a constantly changing world, animals naturally adapt quickly to new environments and adjust their behavior to maximize the benefits. This involves making decisions that will lead to maximum subjective benefit, based on the changing relationships between specific events and outcomes, and thus requires knowing the current worth of each option when making a choice. Typically, an item's worth is learned through direct experience, a process often explained through the concept of classical reinforcement learning(1). However, animals with highly developed brains, particularly primates, have also evolved the ability to infer the value of unexperienced events/items from their knowledge of the world. This ability is described by the rule or theory of shifting relationships. For example, if a monkey is eating a banana

and notices that it is ripe, it may be able to infer that other nearby bananas are also ripe, based on its knowledge of how fruits ripen. Optimal decision-making relies on a balance between experience- and inference-based behavioral strategies. The primate orbitofrontal cortex (OFC) is thought to contribute to such adaptive behavior by leveraging both direct experience and knowledge of the current context. The OFC has long been thought to play an essential role in encoding the subjective values of alternative events/items that guide subsequent decision-making, and in learning/updating these values by integrating past experiences as consequences of our choices(2-6). At the same time, the OFC has also been shown to be necessary for inferring value based on mental simulation of outcomes, even in the absence of direct experience, by generalizing knowledge of the current situation or environment(7-9). A recent report suggests that the OFC regulates the balance between these two valuation strategies, rather than simply initiating one or the other(10). Thus, there is ongoing debate about the core function of the OFC in adaptive behavior.

The complexity of OFC function might arise from its interactions with other brain regions through direct anatomical connections(11). For example, subcortical structures, such as the rostromedial part of the caudate nucleus (rmCD) and the medial part of the mediodorsal thalamus (MDm), receive direct projections from the OFC(12-14). Lesions to these areas have been shown to produce deficits that are similar, but not identical, to those produced by OFC lesions(15-20) with a tendency that rmCD and MDm are particularly involved in value-updating based on experience and inference, respectively, suggesting that they have overlapping yet distinct roles in adaptive behavior. These findings have raised the possibility that two pathways originating from OFC, namely the OFC-rmCD and OFC-MDm pathways, are needed for different value-updating strategies. To investigate this possibility, it is essential to independently manipulate the prefronto-subcortical circuits, which is technically challenging, especially in behaving nonhuman primates.

To investigate the causal roles of the OFC and its originating pathways in these two types of valuation, we use a chemogenetic tool called designer receptors exclusively activated by designer drugs (DREADDs)(21). This tool allows neurons to



**Fig. 1.** Experience- and inference-based value updating in multi-reward reversal learning tasks. (A) Sequence of a trial (left) and the reversal rule for stimulus-reward associations (right). S1-S5 represent the identity of each stimulus and R1-R5 represent the amount of reward (1 to 5 drops of juice) associated with each stimulus. (B and C) Examples of stimulus sets and baseline performance for the “NOVEL” (B) and “FAMILIAR” (C) tasks. Data were averaged for two monkeys ( $N = 16$  and 14 sessions; 8 and 7 sessions for each monkey in NOVEL and FAMILIAR tasks, respectively). Solid lines and shaded area represent the mean and s.e.m., respectively. Data for the FAMILIAR task were truncated to show those around the reversals and were averaged across reversals. (D and E) Bayesian posterior probability calculated for “EXP” and “INF” models given the behavioral data in each task (left) and the behavioral simulation by the model with higher posterior probability (right); EXP model for the NOVEL task (D) and INF model for FAMILIAR task (E), respectively.

be silenced by activating an inhibitory DREADD (hM4Di) following systemic administration of a DREADD agonist. Additionally, local agonist infusion that activates hM4Di expressed at axon terminals can suppress synaptic transmission(22, 23). By combining these techniques with positron emission tomography (PET) as an *in vivo* imaging tool for hM4Di-positive projection sites, we have previously developed imaging-guided chemogenetic synaptic-silencing that is dramatically more efficient and accurate, especially when applied to nonhuman primates(24). Leveraging this technique and a model-fitting analysis in a reinforcement learning framework, the present study addresses the contributions of these two OFC-subcortical pathways to different value-updating strategies. Our results suggest that experience- and inference-based strategies for updating stimulus-reward associations rely on the OFC-rmCD and OFC-MDm pathways, respectively.

## Results

### Experience- and inference-based behavioral adaptation in multi-reward value-updating tasks

To address the question of how the OFC and its projections to subcortical structures contribute to behavioral adaptation through experience- and inference-based valuation, we devised two behavioral tasks for macaque monkeys: NOVEL and FAMILIAR tasks, respectively (Fig. 1A). In both tasks, the monkeys were required to choose either of two presented visual stimuli (out of a set of five abstract images), each of which was associated with 1, 2, 3, 4, or 5 drops of juice.

The order of associations was reversed within a session (Fig. 1A). To maximize their reward, monkeys had to learn the values of the visual stimuli and then update them following subsequent changes in stimulus-reward associations. In the NOVEL task, which was aimed to assess experience-based updating, a new set of stimuli was introduced in each session, thus requiring the monkeys to learn new stimulus-reward associations as well as the association reversal that was imposed mid-session (90 trials after the beginning of each session). After several months of training, two monkeys (Mks #1 and 2) were able to learn new stimulus-reward associations within 80-90 trials and adapted to their reversal within 30-50 trials (Fig. 1B, right). In the FAMILIAR task, which was aimed to assess inference-based updating, a fixed set of five visual stimuli was used throughout the experiments, and the reversals were imposed several times after performance reached a predetermined criterion (see Methods). Following several months of training on this task, the monkeys were able to adapt to the reversal within 3-5 trials (Fig. 1C, right). They even showed optimal choice for “unexperienced” stimulus-reward associations after experiencing the other associations following the reversal (for details, see latter section), indicating that they solved this task based on inference, that is, their prior knowledge of the limited patterns of stimulus-reward associations.

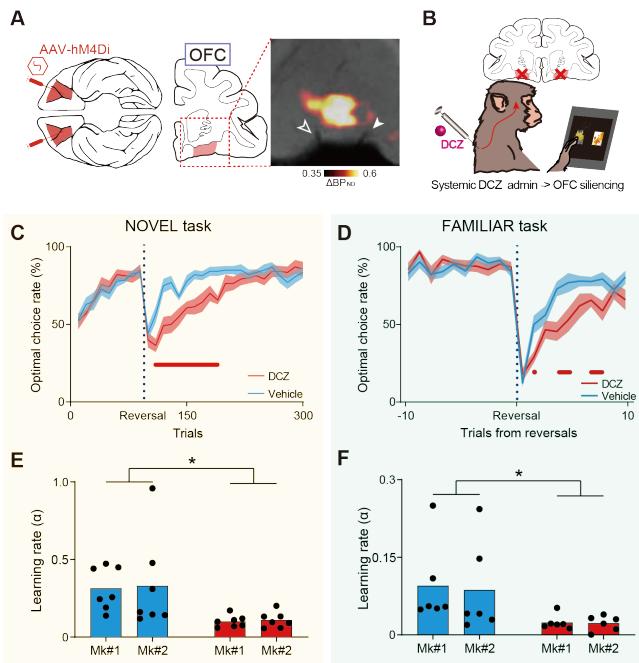
To examine whether the monkeys solved these tasks based on experience- or inference-based strategies, we conducted simulations using two types of reinforcement learning models (see Methods for details); one assuming that the values were updated through direct experience based on standard model-free reinforcement learning that was driven by reward prediction errors (“EXP” model), and one assuming that the monkeys had a priori knowledge of the two possible stimulus-reward association patterns (1,2,3,4,5 or 5,4,3,2,1 drops) through daily training, and that prediction errors drove the transition between these two already learned value sets, allowing them to infer the values of any unexperienced stimuli (“INF” model). As expected, this analysis revealed that behavior during the NOVEL task was better explained by the EXP model (Fig. 1D), while that during the FAMILIAR task was better explained by the INF model (Fig. 1E). These results suggest that monkeys “solved” the two tasks with different strategies — experience-based for the NOVEL task and inference-based for the FAMILIAR task.

### OFC silencing impairs both experience- and inference-based strategies

Next, we chemogenetically inactivated the OFC to determine whether it contributes to experience- and/or inference-based valuation strategies. First, we introduced the inhibitory DREADD, hM4Di, bilaterally into the lateral OFC (Brodmann’s area 11/13) of the two monkeys via injections of an adeno-associated virus (AAV) vector (AAV2-CMV-hM4Di and AAV2.1-CaMKII-hM4Di-IRES-AcGFP for Mks #1 and 2, respectively) (Fig. 2A). Several weeks after the injections, we non-invasively visualized hM4Di expression using PET imaging with DREADD-PET tracers. In both

monkeys, we consistently observed increased PET signal in the bilateral OFC (Fig. 2A; fig. S1A), which was confirmed by post-mortem immunohistochemistry to reflect hM4Di expression. We then examined the effect that silencing the OFC had on behavior; task performances were compared after systemically administering either a vehicle control or the DREADD agonist deschloroclozapine (DCZ) (Fig. 2B). While silencing the OFC did not alter performance on the NOVEL task before reversal of the stimulus-reward contingencies (acquisition phase), it did impair performance following the reversal. This was particularly true in the early phase (Fig. 2C) and was consistently observed in both monkeys (fig. S2A,B). Similarly, OFC silencing also impaired performance on the FAMILIAR task just after reversals (Fig. 2D, fig. S2C,D). To quantify the silencing effect in both tasks, we fitted two behavioral models to the behavioral data following DCZ and vehicle administration. Model-fitting analysis revealed that the impaired performance after OFC silencing could be attributed to a decrease in the learning rate ( $\alpha$ ) after the reversals (Fig. 2E,F, see Methods for details), but not to the change in the extent of exploration (the inverse temperature,  $\beta$ ; fig. S3A,D), suggesting that silencing the OFC led to deficits in value-updating when using either strategy.

We conducted four additional experiments to confirm that the observed behavioral changes were due to the loss of normal OFC function that is involved in adapting to the shift in the task context, and which requires updating the values of the external stimuli. First, to confirm that healthy OFC function is required at the time of the reversal, but not when learning the stimulus-reward associations or during any other process in the acquisition phase, we administered DCZ after the acquisition phase (just before reversal, see Methods for details). This manipulation produced deficits similar to those observed when DCZ was administered before the beginning of each session (fig. S4A), suggesting that the effects of OFC silencing were limited to the reversal of stimulus-reward associations. Second, we confirmed that DCZ alone did not significantly affect behavioral performance before the introduction of hM4Di (Mk #2, fig. S4B), indicating that the effects observed after DCZ administration were due to DREADD activation. Third, we confirmed that OFC silencing did not significantly impact simple reversal learning (fig. S5), consistent with a previous lesion study(25). Fourth, to test whether the OFC is also essential for situations in which knowledge-based value-updating is required and the cognitive load is high, but the item values are binary (i.e., reward or no reward), we examined the effects of OFC silencing using an analog of the Wisconsin Card Sorting Test(26) (Mk #2, fig. S6B). In this case, we found that silencing the OFC had no effect on task performance (fig. S6C), suggesting that it is required specifically when updating external items with complex, non-binary values. Importantly, chemogenetic silencing of the OFC remained effective at the end of all experiments, including these control experiments and the pathway-selective manipulations (see below) (fig. S6A), as demonstrated by the significant effects of DCZ administra-



**Fig. 2.** Chemogenetic silencing of the bilateral OFC impaired experience- and inference-based value updating. (A) Injection area (bilateral OFC, Brodmann's area 11/13) and PET image showing hM4Di expression in MK2. (B) Schema of OFC silencing per se. DCZ (100  $\mu$ g/kg) was systemically injected intramuscularly. (C and D) Behavioral effects of chemogenetic OFC silencing on NOVEL (C) and FAMILIAR (D) task performance. Data for vehicle injections (cyan) and DCZ injections (red) are shown. The optimal choice rate ( $n = 7$  and 6 sessions for each treatment in each monkey for the NOVEL and FAMILIAR tasks, respectively) following DCZ (red) and vehicle injections (cyan) was averaged across two monkeys. The red marks below each graph indicate trial numbers after reversal for which significant differences between vehicle and DCZ conditions were observed ( $t$ -test,  $p < 0.05$ ). (E and F) Estimated learning rates in the NOVEL task with the EXP model (E, two-way ANOVA, treatment  $\times$  monkey, main effect of treatment,  $F(1,24) = 11.3$ ,  $p = 2.6 \times 10^{-3}$ ; subject,  $F(1,24) = 0.04$ ,  $p = 0.84$ ; interaction,  $F(1,24) = 0.004$ ,  $p = 0.95$ ) and in the FAMILIAR task with the INF model (F, treatment,  $F(1,20) = 7.5$ ,  $p = 1.3 \times 10^{-2}$ ; subject,  $F(1,20) = 0.04$ ,  $p = 0.85$ ; interaction,  $F(1,20) = 0.02$ ,  $p = 0.90$ ). Asterisks:  $p < 0.05$  for significant main effect of treatment.

tion on performance during the final devaluation task (fig. S6D,E), which is one of the most common tasks requiring normal OFC function(7, 27). Taken together, these results suggest that the OFC is essential for adapting behavioral responses that are specifically contingent upon being able to update multiple values of multiple external stimuli.

### The OFC-rmCD and OFC-MDm pathways are necessary for experience- and inference-based value-updating, respectively

Having demonstrated that the OFC is essential for both experience- and inference-based behavioral adaptation, the next question is whether these different strategies are governed by separate neural pathways. To answer this, we conducted a chemogenetic pathway-selective manipulation wherein information flow can be temporarily inactivated by local agonist infusion into axonal terminals expressing hM4Di. It is generally challenging to precisely localize and target axonal projection sites *in vivo* in monkeys, which have relatively large and complexly shaped brains. We overcame this obstacle by using an imaging-guided chemogenetic technique(24) in which PET allows localization of the pro-

jection sites for hM4Di-positive OFC neurons. Aside from the PFC, subtraction PET images (post-AAV injection minus pre-AAV injection) showed increased PET signals in the striatum and the thalamus, specifically in the rmCD, MDm, and the medial part of the putamen (Fig. 3A,B; fig. S1B-D). These regions colocalized with GFP-positive axon terminals under immunohistological examination (fig. S1B-D). In contrast, the hM4Di signal was not clearly observed via PET or histology in other brain regions which are known to also receive projections from the OFC, such as the amygdala, and was not comparable to what we observed in these three regions (fig. S1E). Thus, our next experiments focused on the projections from the OFC to the three DREADD-positive terminal regions.

To reversibly block neuronal transmission from the OFC, we infused DCZ into one of the three terminal regions under the guidance of MR, PET, and CT imaging (Fig. 3A,B). Compared with the control vehicle infusion, local DCZ infusion into the rmCD impaired performance on the NOVEL task just after reversals, similar to the systemic DCZ injections (Fig. 3C, left). However, this was not the case for the FAMILIAR task (Fig. 3C, right). Conversely, DCZ infusion into MDm resulted in relatively minor impairment on the NOVEL task (Fig. 3D, left), but significant impairment on the FAMILIAR task (Fig. 3D, right). Although we also injected DCZ into the medial part of the putamen, where we found weak PET and histological signals (fig. S1D), this did not affect performance on the NOVEL task (fig. S7). We therefore focused further tests on the OFC to rmCD and OFC to MDm pathways. Model fitting analysis using two reinforcement learning models revealed that silencing the OFC-rmCD pathway significantly reduced the learning rate during the NOVEL task (EXP model; Fig. 3E, left), but not during the FAMILIAR task (INF model; Fig. 3E, right). Conversely, silencing the OFC-MDm pathway had no impact on the learning rate during the NOVEL task (Exp model; Fig. 3F, left), but significantly reduced it during the FAMILIAR task (INF model; Fig. 3F, right). Similar to OFC silencing, silencing either pathway did not affect the inverse temperature (Extended Data Fig. 3b-f).

Damage to the OFC in humans has been associated with increased impulsivity(28). In monkeys, OFC inactivation or lesioning has resulted in faster reaction times on experimental tasks(29), which is generally interpreted as a sign of impulsivity or lack of control. To assess whether silencing the OFC and its projections affected impulsivity in our task context, and if so, whether this behavioral change is related to the observed impairment in performance, we examined reaction times after each type of silencing (fig. S8). OFC silencing resulted in shorter reaction times for both monkeys during all task phases, but we did not observe any direct relationship between reaction time and performance (Extended Data Fig. 8a). In contrast, silencing the OFC-rmCD and OFC-MDm pathways induced complex and contradicting results; silencing the OFC-rmCD pathway increased reaction time (fig. S8B), whereas silencing the OFC-MDm pathway did not influence reaction time (fig. S8C). Although these

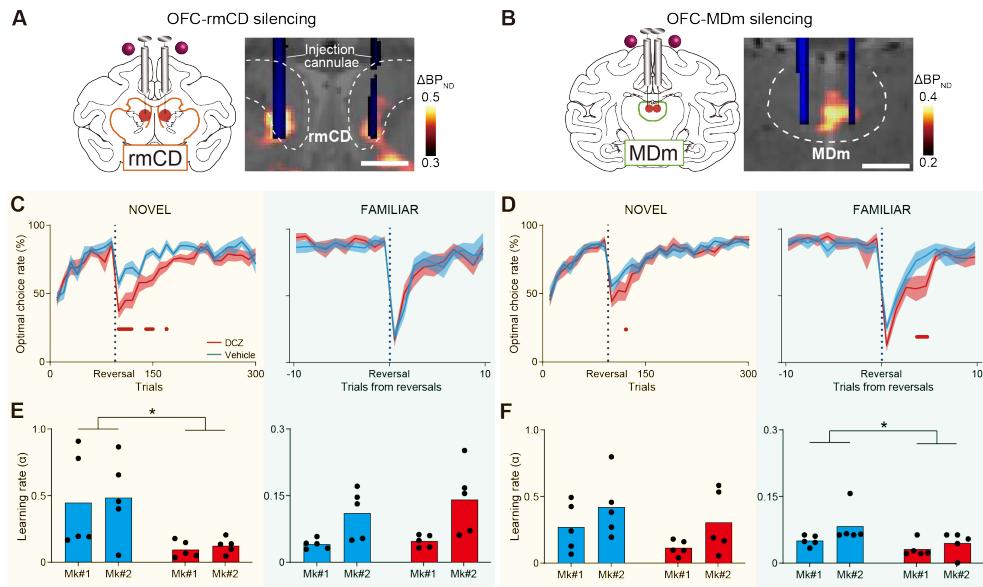
effects on reaction time differed, we can conclude that the difficulty in updating item values was not related to increased impulsivity.

Taken together, selective silencing of the two OFC projections indicated that both are involved in updating stimulus values; the OFC-rmCD pathway is needed when updating based on direct experience of stimulus-reward associations, whereas the OFC-MDm pathway is needed when updating based on inference from previously learned knowledge.

### The OFC-rmCD and OFC-MDm pathways differentially contribute to the sensitivity to past outcomes

Our results suggest that the two pathways contribute distinctly to different behavioral strategies. To further corroborate this dissociation, we perform a learning-model agnostic analysis in which we decompose trials into specific events related to each behavioral strategy. First, we focused on experience-based updating during the NOVEL task. If the ability to update values based on past experience is impaired, the behavior following unpredicted positive experiences (i.e., obtaining a good result after choosing the previously unchosen option) should favor repeating the same choice, and vice versa following negative experiences (Fig. 4A). Consistent with the model-fitting analysis, silencing the OFC or the OFC-rmCD pathway significantly impaired performance following both positive and negative experiences (Fig. 4C,D), suggesting that their contribution to updating stimulus-reward associations is based on both positive and negative experiences. In contrast, silencing the OFC-MDm pathway only impaired performance following negative experiences (Fig. 4E), likely reflecting minor deficits at the immediate post-reversal period (Fig. 3D). Similarly, asymmetric deficit was induced by silencing the OFC-MDm pathway during the FAMILIAR task (fig. S9).

Next, we focused on inference-based updating during the FAMILIAR task. We found that the monkeys could adapt to the new stimulus-reward association after a single experience of a reversal (1st trial, Fig. 4B, left), even when the subsequent trial did not include options that appeared in the previous trial (inference trial, Fig. 4B, right). Monkeys exhibited a greater percentage of optimal choices in the inference trial than in the 1st trial (inference vs. first trials; Mk #1: 72.2% vs 19.6%; Mk #2: 59.3% vs 17.7%) under the baseline control conditions, indicating that they inferred the stimulus-reward associations without needing direct experience. This inference-based behavioral adaptation was significantly impaired following silencing of the OFC or the OFC-MDm pathway (Fig. 4F,H). In contrast, silencing the OFC-rmCD pathway had no effect on the inference trials (Fig. 4G). Taken together, these results support the conclusion that the OFC-rmCD pathway is essential for updating value via positive and negative experiences, whereas the OFC-MDm pathway is selectively involved in behavioral changes following negative experiences—a capacity that is critical for rapid behavioral adaptation using an inference-based strategy based on prior knowledge of the situation.



**Fig. 3.** OFC-rmCD and OFC-MDm pathways are necessary for experience- and inference-based value-updating, respectively. (A and B) vChemogenetic silencing of the OFC-rmCD (A) and OFC-MDm (B) pathways by local DCZ infusion into either bi-lateral rmCD or MDm, specifically at hM4Di-positive OFC terminal sites. A CT image showing the infusion cannulae (blue) overlaying a structural MR image (grey), and a PET image showing a high [11C]DCZ binding region (hM4Di expression, hot color) obtained from Mk2. The dashed lines represent the borders of the caudate nucleus and mediodorsal thalamus, respectively. (C and D) Optimal choice rate ( $n = 10$  sessions for each treatment, 5 sessions in each monkey) in the NOVEL (left) and FAMILIAR (right) tasks after silencing the OFC-rmCD (C) and OFC-MDm (D) pathways. Conventions are the same as in Fig. 2. (E and F) Estimated learning rates after silencing the OFC-rmCD pathway during the NOVEL (E, left, treatment,  $F(1,16) = 11.0$ ,  $p = 4.4 \times 10^{-3}$ ; subject,  $F(1,16) = 0.10$ ,  $p = 0.76$ ; interaction,  $F(1,16) = 0.0022$ ,  $p = 0.96$ ) and FAMILAR (E, right, treatment,  $F(1,16) = 0.74$ ,  $p = 0.40$ ; subject,  $F(1,16) = 14.1$ ,  $p = 1.8 \times 10^{-3}$ ; interaction,  $F(1,16) = 0.30$ ,  $p = 0.59$ ) tasks, and the OFC-MDm pathway during the NOVEL (F, left, treatment,  $F(1,16) = 2.5$ ,  $p = 0.14$ ; subject,  $F(1,16) = 3.9$ ,  $p = 0.07$ ; interaction,  $F(1,16) = 0.059$ ,  $p = 0.81$ ) and FAMILAR (F, right, treatment,  $F(1,16) = 5.7$ ,  $p = 2.9 \times 10^{-2}$ ; subject,  $F(1,16) = 3.7$ ,  $p = 0.07$ ; interaction,  $F(1,16) = 0.61$ ,  $p = 0.45$ ). Scale bars: 5 mm.

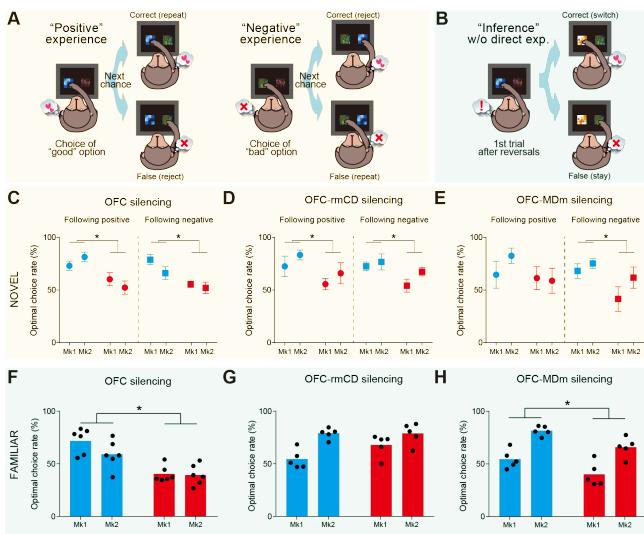
## Discussion

Here, by combining chemogenetic silencing of individual neural pathways with a model-fitting approach, we demonstrate that the two natural strategies for updating subjective value of external stimuli rely on two distinct neural pathways, each originating in the OFC and projecting to a different subcortical brain regions. Silencing the OFC-rmCD pathway impaired performance when monkeys updated option values through direct experience of both positive and negative changes of stimulus-reward associations, whereas silencing the OFC-MDm pathway impaired performance when monkeys updated the values based on inference that is guided by negative experience which could notice the change of the task context. This dissociable contribution of neural pathways provides new insights into the neural basis of value-based adaptive decision-making in primates.

The OFC is widely thought to play a central role in updating and maintaining information about possible outcomes(6, 11). Both lesion and recording studies have shown that the OFC is critical for updating associations between stimulus/action and outcome, but not for acquiring such associations(20, 30), which was also the case in our study (Fig. 2). Previous research has also suggested that the OFC is required for reversal learning in various species(31-33). However, recent studies have challenged this view by showing that selective lesion/inactivation of the OFC in macaque monkeys had no effect in a simple reversal learning paradigm(25, 34), a result consistent with our observations (fig. S5). Recently, another study suggested that the OFC is essential for updating the desirability (i.e., quality or quantity) of reward-associated stimuli, but not their availability (i.e., the probability of

receiving rewards)(20). Our current results, in which OFC silencing severely impaired performance on a behavioral task that requires monkeys to associate multiple stimuli with multiple reward amounts and to update their associations, support this view. Notably, OFC silencing did not impact performance on the Wisconsin Card Sorting Task (WCST, fig. S6), which is commonly used to assess behavioral flexibility. Although previous lesion studies in monkeys have reported deficits in WCST performance following aspirating lesions(20), subsequent studies have suggested that these deficits may have been due to damaged fibers that pass near the OFC(25). Because outcomes on the WCST are typically binary, either receiving a reward or not, our results suggest that primate OFC, at least the lateral part that we focused on in this study (Brodmann's area 11/13), is specialized for representing stimulus-reward associations that are based on the desirability of possible outcomes.

Recent research into the role of the OFC has highlighted a specialized role in inference-based, or model-based, learning. This idea has been supported by a number of recording and lesion studies, which have consistently demonstrated the involvement of the OFC in such learning processes(35, 36). However, other recent reports suggest that rodent OFC, or the ventral PFC in humans, is not simply engaged in either experience- or inference-based value-updating strategies, but rather in regulating the switch between them according to current contexts(10, 37). Our present results are consistent with and extend this idea; we can now assign distinct roles for at least two different OFC projections to subcortical structures (rmCD and MDm). To the best of our knowledge, the present study is the first report in primates that has



**Fig. 4.** Silencing of OFC-rmCD and OFC-MDm pathways differentially affected the sensitivity to past outcomes. (A and B) Schematic drawings showing the trials following positive (left) and negative (right) experience (A), and the “inference” trial following a single experience of the reversal of stimulus-reward associations (B) for the NOVEL and FAMILIAR tasks, respectively. (C to E) Averaged optimal choice rate for trials following positive (left) and negative (right) outcomes for OFC silencing (C) (positive: treatment,  $F(1,24) = 14.8$ ,  $p = 7.9 \times 10^{-4}$ ; subject,  $F(1,24) = 0.001$ ,  $p = 0.97$ ; interaction,  $F(1,24) = 2.2$ ,  $p = 0.15$ ; negative: treatment,  $F(1,24) = 14.3$ ,  $p = 9.1 \times 10^{-4}$ ; subject,  $F(1,24) = 2.7$ ,  $p = 0.11$ ; interaction,  $F(1,24) = 0.94$ ,  $p = 0.34$ ), OFC-rmCD silencing (D) (positive: treatment,  $F(1,16) = 4.8$ ,  $p = 4.3 \times 10^{-2}$ ; subject,  $F(1,16) = 1.8$ ,  $p = 0.20$ ; interaction,  $F(1,16) = 0.001$ ,  $p = 0.97$ ; negative: treatment,  $F(1,16) = 6.0$ ,  $p = 2.6 \times 10^{-2}$ ; subject,  $F(1,16) = 2.4$ ,  $p = 0.14$ ; interaction,  $F(1,16) = 0.65$ ,  $p = 0.43$ ), and OFC-MDm silencing (E) (positive: treatment,  $F(1,16) = 1.5$ ,  $p = 0.23$ ; subject,  $F(1,16) = 0.51$ ,  $p = 0.48$ ; interaction,  $F(1,16) = 0.90$ ,  $p = 0.36$ ; negative: treatment,  $F(1,16) = 5.2$ ,  $p = 3.7 \times 10^{-2}$ ; subject,  $F(1,16) = 2.4$ ,  $p = 0.14$ ; interaction,  $F(1,16) = 0.56$ ,  $p = 0.46$ ). (F to H) Averaged optimal choice rate for inference trials after OFC silencing (F) (treatment,  $F(1,20) = 33.1$ ,  $p = 1.3 \times 10^{-5}$ ; subject,  $F(1,20) = 2.5$ ,  $p = 0.13$ ; interaction,  $F(1,20) = 1.7$ ,  $p = 0.21$ ), OFC-rmCD silencing (G) (treatment,  $F(1,16) = 2.7$ ,  $p = 0.12$ ; subject,  $F(1,16) = 19.6$ ,  $p = 4.2 \times 10^{-4}$ ; interaction,  $F(1,16) = 2.9$ ,  $p = 0.11$ ), and OFC-MDm silencing (H) (treatment,  $F(1,16) = 14.2$ ,  $p = 1.7 \times 10^{-3}$ ; subject,  $F(1,16) = 43.7$ ,  $p = 6.0 \times 10^{-6}$ ; interaction,  $F(1,16) = 0.021$ ,  $p = 0.88$ ).

directly investigated the dissociable roles of OFC-originating pathways. Importantly, the two pathways we focused on form part of a broader circuit for value-updating. Therefore comprehensively examining the other circuitry, such as the OFC-amygdala pathway(38-40) and the OFC-sensory cortex pathway(41), is essential for a deeper understanding of value-guided behavior.

In the current study, we were able to dissect the functions of the OFC-subcortical circuits by using the imaging-guided pathway-selective synaptic silencing technique that we recently developed(24). Interestingly, in that study, we also observed a functional dissociation in frontal cortex pathways, in that case originating in the dorsolateral prefrontal cortex (dIPFC) and projecting to either the CD or MD thalamus. Although the OFC and dIPFC are located in different parallel cortico-basal ganglia-thalamo-cortical circuits(42), the observed similarity that different subcortical pathways have different functions suggests that the parallel prefrontal-subcortical networks might share a common neural basis for adaptive behavior. Future studies should aim to better understand the overall architecture of these networks by identifying where and how they communicate

with each other and how the information processed in each network is integrated.

In summary, leveraging the technical advantage of imaging-guided pathway-selective chemogenetic silencing, we have demonstrated the dissociable contributions of two OFC-subcortical pathways to different value-updating strategies. The identification of causal relationships between a specific neural pathway and a cognitive function, as demonstrated in this study, can compliment what we have learned from human studies, but which cannot be directly tested in humans for ethical reasons. Thus, in addition to providing new insights into the neural basis of value-based adaptive decision-making in nonhuman primates, our results have implications for psychiatric conditions associated with malfunctioning prefrontal-subcortical networks, such as obsessive-compulsive disorder.

## ACKNOWLEDGEMENTS

Two Japanese monkeys were provided by the Japan MEXT National Bio-Resource Project “Japanese Monkeys”. We thank Drs. Brian Russ and Junya Hirokawa for their valuable comments on an earlier version of the manuscript, and Jun Kamei, Ryuji Yamaguchi, Yuichi Matsuda, Yoshio Sugii, Takashi Okauchi, Rie Yoshida, Maki Fujiwara, and Mayuko Nakano for their technical assistance. This study was supported by MEXT/JSPS KAKENHI Grant Numbers JP18K15353, JP21K07268, JP22H05521 (to KO), and, JP17H02219 (to TH), JP22H05157 (to KI), JP19H05467 (to MT), and JP20H05955 (to TM), by JST PRESTO Grant Numbers JPMJPR22S3 (to KO), and JPMJPR2128 (to K Majima) and by AMED Grant Numbers JP23dm0307007 (to TH), JP21dm0107146 (to TM), JP20dm0307021 (to KI), and JP21dm0207077 (to MT). M.A.G.E and B.J.R were supported by the Intramural Research Program, NIMH/NIH/DHHS (project# ZIAHM 002032).

## Materials & Methods

**Subjects.** Two male Japanese monkeys (*Macaca fuscata*) participated in the experiments (Mk #1: 7.2 kg; Mk #2: 6.8 kg; both aged 4 years at the beginning of experiments). The monkeys were kept in individual primate cages in an air-conditioned room. A standard diet, supplementary fruits/vegetables, and a tablet of vitamin C (200 mg) were provided daily. All experimental procedures involving the monkeys were carried out in accordance with the Guide for the Care and Use of Nonhuman primates in Neuroscience Research (The Japan Neuroscience Society; [https://www.jnss.org/en/animal\\_primates](https://www.jnss.org/en/animal_primates)) and were approved by the Animal Ethics Committee of the National Institutes for Quantum Science and Technology.

**Viral vector production.** Mk #1 was co-injected with two AAV vectors, one expressing hM4Di and the other expressing GFP (AAV2-CMV-hM4Di and AAV2-CMV-AcGFP;  $2.3 \times 10^{13}$  and  $4.6 \times 10^{12}$  particles/mL, respectively). Mk #2 was injected with an AAV vector expressing both hM4Di and GFP (AAV2.1-CaMKII-hM4Di-IRES-AcGFP,  $1.0 \times 10^{13}$  particles/mL). AAV vectors were produced by a helper-free triple transfection procedure, and were purified by affinity chromatography (GE Healthcare, Chicago, USA). Viral titer was determined by quantitative PCR using Taq-Man technology (Life Technologies, Waltham, USA)(43).

**Surgical procedures and viral vector injections.** Surgeries were performed under aseptic conditions in a fully equipped operating suite. We monitored body temperature, heart rate, SpO<sub>2</sub>, and tidal CO<sub>2</sub> throughout all surgical procedures. Monkeys were immobilized by intramuscular (i.m.) injection of ketamine (5–10 mg/kg) and xylazine (0.2–0.5 mg/kg) and intubated with an endotracheal tube. Anesthesia was maintained with isoflurane (1%–3%, to effect). Before surgery, magnetic resonance (MR) imaging (7 tesla 400 mm/SS system, NIRS/KOBELCO/Brucker) and X-ray computed tomography (CT) scans (Accuitomo170, J. MORITA CO., Kyoto, Japan) were performed under anesthesia (continuous intravenous infusion

of propofol 0.2–0.6 mg/kg/min). Overlay MR and CT images were created using PMOD® image-analysis software (PMOD Technologies Ltd, Zurich, Switzerland) to estimate stereotaxic coordinates of target brain structures. After surgery, prophylactic antibiotics and analgesics (cefmetazole, 25–50 mg/kg; ketoprofen, 1–2 mg/kg) were administered.

The bilateral OFCs (BA11 & BA13) of each monkey were injected with the AAV vectors (Fig. 1A). The injections were performed under direct vision using the same types of surgical procedures as in a previous study(44). Briefly, after retracting skin, galea, and muscle, the frontal cortex was exposed by removing a bone flap and reflecting the dura mater. Then, handheld injections were made under visual guidance through an operating microscope (Leica M220, Leica Microsystems GmbH, Wetzlar, Germany), with care taken to place the beveled tip of a microsyringe (Model 1701RN, Hamilton) containing the viral vector at an angle oblique to the brain surface. The needle (26 Gauge, PT2) was inserted into the intended area of injection by one person and a second person pressed the plunge #1 and MK #2 via 53 and 49 tracks, respectively.

**PET imaging.** PET imaging was conducted as previously reported(3). Briefly, PET scans were conducted before injection of vectors and at 45 days after injection for both monkeys. PET scans were performed using a microPET Focus 220 scanner (Siemens Medical Solutions USA, Malvern, USA). Monkeys were immobilized by ketamine (5–10 mg/kg) and xylazine (0.2–0.5 mg/kg) and then maintained under anesthetized condition with isoflurane (1%–3%) during all PET procedures. Transmission scans were performed for approximately 20 min with a Ge-68 source. Emission scans were acquired in 3D list mode with an energy window of 350–750 keV after intravenous bolus injection of [<sup>11</sup>C]clozapine (for Mk1; 375.5–394.7 MBq) or [<sup>11</sup>C]DCZ (for MK2; 324.9–382.3 MBq). Emission data acquisition lasted 90 min. To estimate the specific binding of [<sup>11</sup>C]DCZ in Mk2, regional binding potential relative to nondisplaceable radioligand (BPND) was calculated by PMOD® with an original multilinear reference tissue model (MRTMo). To visualize the expression of DREADDs, contrast (subtraction) of images taken before and 45 days after vector injection were created using PMOD for SUV (standardized uptake value) for Mk1 and BPND for Mk2 by investigating whether differential PET signals were observed at the target sites.

**Drug administration.** DCZ (HY-42110; MedChemExpress) was dissolved in 2.5% dimethyl sulfoxide (DMSO, FUJIFILM Wako Pure Chemical Co.), aliquoted and stored at –30°C. For systemic i.m. injection, this stock solution was first diluted in saline to a final volume of 100 µg/kg. Fresh solution was prepared on each day of usage.

For microinfusion, DCZ was first dissolved in DMSO and then diluted in PBS to a final concentration of 100 nM. We prepared fresh solutions on the day of usage. We used two stainless steel infusion cannulae (outer diameter 300 µm; Muromachi-Kikai) inserted into each target region: rmCD and MDm, and ventral putamen for additional experiments (Fig. S7). Each cannula was connected to a 10-µL microsyringe (7105KH; Hamilton) via polyethylene tubing. These cannulae were advanced via guide tube by means of an oil-drive micromanipulator. DCZ solution or PBS was injected at a rate of 0.25 µL/min by auto-injector (Legato210; KD Scientific) for a total volume of 3 µL for each hemisphere. The injection volumes were determined based on a previous study reporting that injections of 3 µL and 1.5 µL resulted in a diameter of aqueous spread in the monkey brain of approximately 5–6 mm and 3–4 mm, respectively(45). We chose sufficient volumes to cover the hM4Di-positive terminal sites, which had diameters of 5–7 mm and 3–4 mm for the rmCD and MDm, respectively, as measured by increased PET signals. Because the MDm is located close to the midline, we placed the cannulae laterally near the MDm so that the injected solution would diffuse into the entirety of the MDm (Fig. 3B). CT image was obtained to visualize the infusion cannulae in relation to the chambers and skull following each infusion. The CT image was overlaid on MR and PET images obtained beforehand using PMOD to verify that the infusion sites (tips of the infusion cannulae) were located in the target (presumed hM4Di-positive terminal regions identified as increased PET signals). The behavioral session began approximately 30 min after the end of the infusion and lasted approximately one hour. We performed at most one silencing experiment per week for one area.

**Behavioral tasks.** The monkeys were tested with two versions of modified reversal learning tasks in which they were required to choose either of two visual stimuli (out of a set of five) presented on a computer screen. Behavioral testing was conducted in a sound-attenuated room. The monkeys sat on a monkey chair from which they could reach out one with hand to touch an LCD display placed in front of them. The behavioral task was controlled by a computer using commercially available software (Inquisit, Millisecond). A monkey initiated a trial by touching a sensor mounted on the chair, which caused a small white circle to appear in the center of the display. After a delay of 0.5 s, the circle disappeared and two stimuli of the five possible stimuli were presented simultaneously on the left and right side of the display. If the monkey touched either stimulus, it could receive a reward from the spout placed in front of its mouth. If the monkey released the touch lever before the presentation of visual stimuli, the trial was aborted and repeated after a 3–4 s inter-trial interval. Each stimulus was associated with 1, 2, 3, 4, or 5 drops of juice.

In the NOVEL task, a new set of visual stimuli was introduced each session, which required the monkeys to learn a new set of stimulus-reward associations. A daily session consisted of 90 acquisition-phase trials, followed by the reversal of the stimulus-reward associations, and then 210 post-reversal trials. The reversal was conducted such that a stimulus previously associated with 1 drop of juice became associated with 5 drops of juice, and one associated with 2 drops became associated with 4 drops, and vice versa. The combination of visual stimuli seen on each trial was pre-determined pseudorandomly so that each combination appeared once every 10 trials in a round-robin fashion.

In the FAMILIAR task, a fixed sets of five visual stimuli were used throughout the experiments. This ensured that the monkeys became familiar with all possible stimulus-reward associations for the two sets before and after reversals. If the optimal choice rate (i.e., the proportion of trials in which the option associated with the greater reward was chosen) in 30 consecutive trials passed 76%, the associations were reversed. Daily sessions comprised 300 and 400 trials for Mk1 and Mk2, respectively. Both monkeys were trained on the NOVEL task and then the FAMILIAR task. As a control, the monkeys were tested on a simple reversal learning task (Extended Data Fig. 5) in which only two novel stimuli were introduced in each session. The stimuli were associated with 5 drops of juice reward or no reward.

One monkey (Mk2) was also tested on a reinforcer devaluation task and on a modified version of the Wisconsin Card Sorting task, as previously described<sup>20,24</sup>. Briefly, in the reinforcer devaluation task, the monkey was required to choose one of two objects that were placed above two holes located in a wooden plate. For one set of objects, food 1 (peanut) was delivered when the object was selected, whereas for the other set of objects, food 2 (raisin) was delivered. The associations between the objects and reward type were fixed throughout training. The monkey was tested on 4 consecutive days in a week: review of object-reward associations (Day 1), baseline choice test (Day 2), review of object-reward associations (Day 3), and choice test following selective satiation (devaluation of a food) with vehicle or DCZ administration 15 min before devaluation (Day 4). The monkey's ability to adaptively shift away from choosing objects associated with the devalued food was calculated as the “proportion shifted” as below, Proportion shifted =  $(F1N-F1D)+(F2N-F2D))/((F1N+F2N))$  where F1 and F2 represent choices associated with the two food types (peanut and raisin) in each week in which that food type was devalued, and D and N respectively represent the data for devaluation (Day 4) and baseline (Day 2).

**Statistics.** Throughout the manuscript, we compared learning rate and optimal choice rate for each monkey between conditions (vehicle and DCZ injections). Comparisons were analyzed with two-way repeated-measures ANOVA (treatment × monkey) to examine the effect of each treatment and individual differences. The analyses were conducted using GraphPad Prism 9. For the NOVEL task, the optimal choice rate was averaged across 10 trials of pseudo-random stimulus-reward combinations (see “Behavioral task”) for each session. In the FAMILIAR task, the optimal choice rate around the time of reversal was averaged across each reversal for each session. Reaction time in the NOVEL task was defined time between releasing the bar and touching the object on the screen, and data were averaged across 10 trials as described above. For statistical analysis of the post-reversal behavioral performance on the NOVEL task, we used the

average optimal choice rate 100, 100, and 50 trials after the reversal for the OFC-, OFC-rmCD, and OFC-MDm silencing conditions, respectively. For the FAMILIAR task, we used the averages from the 10, 10, and 5 trials after the reversal, respectively. These trial numbers roughly corresponded to when behavioral deficits were observed if any.

**Histology and immunostaining.** For histological inspection, monkeys were deeply anesthetized with an overdose of sodium pentobarbital (80 mg/kg, i.v.) and transcardially perfused with saline at 4°C, followed by 4% paraformaldehyde in 0.1 M phosphate buffered saline (PBS), pH 7.4. The brain was removed from the skull, postfixed in the same fresh fixative overnight, saturated with 30% sucrose in phosphate buffer (PB) at 4°C, and then cut serially into 50-μm-thick sections with a freezing microtome. For visualization of immunoreactive GFP signals (co-expressed with hM4Di), a series of every 6th section was immersed in 1% skim milk for 1 h at room temperature and incubated overnight at 4°C with rabbit anti-GFP monoclonal antibody (1:500, G10362, Thermo Fisher Scientific) in PBS containing 0.1% Triton X-100 and 1% normal goat serum for 2 days at 4°C. The sections were then incubated in the same fresh medium containing biotinylated goat anti-rabbit IgG antibody (1:1,000; Jackson ImmunoResearch, West Grove, PA, USA) for 2 h at room temperature, followed by avidin-biotin-peroxidase complex (ABC Elite, Vector Laboratories, Burlingame, CA, USA) for 2 h at room temperature. For visualizing the antigen, the sections were reacted in 0.05 M Tris-HCl buffer (pH 7.6) containing 0.04% diaminobenzidine (DAB), 0.04% NiCl<sub>2</sub>, and 0.003% H<sub>2</sub>O<sub>2</sub>. The sections were mounted on gelatin-coated glass slides, air-dried, and cover-slipped. A portion of the other sections was Nissl-stained with 1% Cresyl violet. Images of sections were digitally captured using an optical microscope equipped with a high-grade charge-coupled device (CCD) camera (Biorevo, Keyence, Osaka, Japan).

**Model-fitting analysis.** We constructed two mathematical models termed “EXP” and “INF” to investigate whether monkey behavior could be explained by experience- or inference-based strategies. In the EXP model, the probability that the learner (i.e., monkey) chooses stimulus  $S_i$  when stimuli  $S_i$  and  $S_j$  are presented ( $i, j = 1, \dots, 5$ ) is given by

$$Pr[S=S_i] = \exp((\beta V_i) / (\exp((\beta V_i) + \exp((\beta V_j))),$$

where  $\beta$  is a parameter controlling the exploration-exploitation trade-off (inverse temperature), and  $V_i$  is the subjective value of stimulus  $S_i$ , which is interpreted as the amount of reward expected by the learner after choosing  $S_i$ .  $V_i$  is updated after each trial based on the Rescorla-Wagner rule, which applies the stochastic gradient-descent algorithm to minimize the squared error between the expected reward for the chosen stimulus and the experienced reward. The update rule after choosing stimulus  $S_i$  is given by

$$V_i \leftarrow V_i - \alpha(V_i - r_{obs}),$$

where  $\alpha$  is the learning rate and  $r_{obs}$  is the amount of the reward provided in the current trial. All  $V_i$  ( $i = 1, \dots, 5$ ) were assumed to be zero at the beginning of each session, and the learning rate and inverse temperature were fitted to the behavioral data in each session via maximum likelihood estimation. In the INF model, we assume that the learner knows that the stimulus-reward association takes one of the two possible patterns. Taking this assumption, we prepared two sets of  $V_i$ :

$$V_i^A = i \quad (i = 1, 2, 3, 4, 5),$$

for pattern A, and

$$V_i^B = 6 - i \quad (i = 1, 2, 3, 4, 5),$$

for pattern B. In the INF model, the learner’s choice is assumed to follow a mixture of the two softmax distributions corresponding to the two possible stimulus-reward association patterns. The probability that stimulus  $S_i$  is chosen when stimuli  $S_i$  and  $S_j$  are presented is given by

$$Pr[S=S_i] = w \{ \exp((\beta V_i^A) / (\exp((\beta V_i^A) + \exp((\beta V_j^A))) \} + (1-w) \{ \exp((\beta V_i^B) / (\exp((\beta V_i^B) + \exp((\beta V_j^B))) \}$$

where  $\beta$  is the inverse temperature and  $w$  is a weight parameter adaptively learned based on the experienced rewards. If  $w = 1$ , the learner’s choice is assumed to follow the softmax distribution corresponding to stimulus-reward association pattern A, and if  $w = 0$ , the choice is assumed to follow the softmax distribution corresponding to stimulus-reward association pattern B. Thus, this weight parameter can be interpreted as a parameter representing the confidence that the current stimulus-reward association is pattern A out of the two possible patterns. Unlike in the EXP model, in the INF model, the inputs to the softmax function are fixed, and the parameter weighting the two softmax distributions is updated after each trial. Like the EXP model, the weight parameter is updated by the update rule derived from the stochastic gradient-descent algorithm, which minimizes the squared error between the expected reward for the chosen stimulus and the experienced reward. The update rule after choosing stimulus  $S_i$  is given by

$$w \leftarrow w - \alpha \{ (w V_i^A + (1-w) V_i^B) - r_{obs} \} (V_i^A - V_i^B)$$

where  $\alpha$  is the learning rate and  $r_{obs}$  is the observed reward in the current trial.  $w$  is assumed to be 0.5 at the beginning of each session and is assumed to take a value in the range of [0.0 1.0] at each update by replacing its value with 0.0/1.0 if it is smaller/larger than 0.0/1.0. As in the EXP model, the learning rate and inverse temperature were fitted to the behavioral data in each session via maximum likelihood estimation.

Because we were interested in the updating behavior during the post-reversal period, we modified the two models described above for the current study. In the modified models, the trials in each session were divided into two groups: 1) the trials in the post-reversal period, and 2) others. The post-reversal period consisted of all trials after the reversal in the NOVEL task, and the five trials after each reversal in the FAMILIAR task. While the learning rate in the basic models is fixed in each session, in the modified models we assumed that it took a different value each of the two trial groups. The learning rate for the post-reversal period is considered to better reflect the updating behavior, and a total of three parameters (two learning rates and inverse temperature) were fitted to the given behavioral data. Unless stated otherwise, the learning rates for the post-reversal periods derived from the modified models were reported in this paper.

To compare the ability of the EXP and INF models to explain the behavioral data, we performed Bayesian model comparison. Following the procedure in previous studies(46, 47), the Bayesian information criterion (BIC) was computed for each model and each session, then the Bayesian posterior probability for each model was computed based on those BIC values.

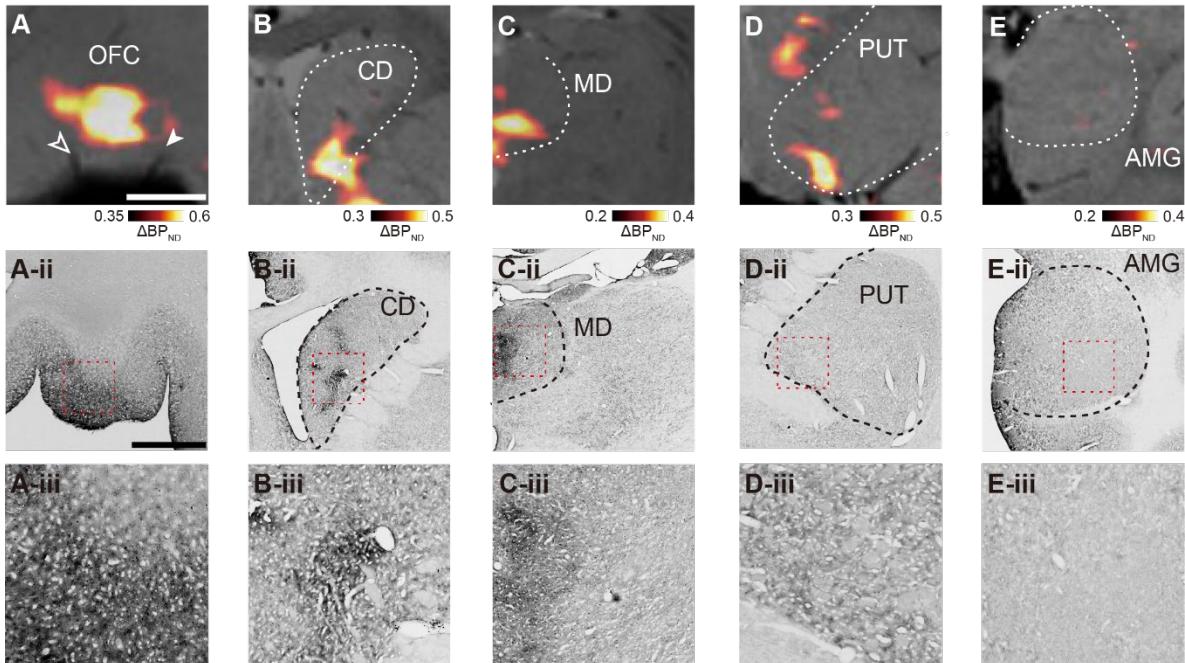
## References

1. R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction. (MIT press, Cambridge, MA, USA, 2018).
2. L. Tremblay, W. Schultz, Relative reward preference in primate orbitofrontal cortex. *Nature* 398, 704-708 (1999).
3. C. Padoa-Schioppa, J. A. Assad, Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223-226 (2006).
4. A. Tsuchida, B. B. Doll, L. K. Fellows, Beyond reversal: A critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *The Journal of Neuroscience* 30, 16868-16875 (2010).
5. K. J. Miller, M. M. Botvinick, C. D. Brody, Value representations in the rodent orbitofrontal cortex drive learning, not choice. *eLife* 11, e64575 (2022).
6. E. T. Rolls, The Orbitofrontal Cortex and Reward. *Cerebral Cortex* 10, 284-294 (2000).
7. T. A. Stalnaker, N. K. Cooch, G. Schoenbaum, What the orbitofrontal cortex does not do. *Nature Neuroscience* 18, 620-627 (2015).
8. T. Kahnt, G. Schoenbaum, Cross-species studies on orbitofrontal control of inference-based behavior. *Behavioral Neuroscience* 135, 109-119 (2021).
9. J. Zhou et al., Rat orbitofrontal ensemble activity contains multiplexed but dissociable representations of value and task structure in an odor sequence task. *Current Biology* 29, 897-907.e893 (2019).
10. M. Panayi, M. Khamassi, S. Killcross, The rodent lateral orbitofrontal cortex as an arbitrator selecting between model-based and model-free learning systems. *Behavioral Neuroscience* 135, 226-244 (2021).
11. P. H. Rudebeck, E. L. Rich, Orbitofrontal cortex. *Current Biology* 28, 1083-1088 (2018).
12. N. R. McFarland, S. N. Haber, Thalamic relay nuclei of the basal

ganglia form both reciprocal and nonreciprocal cortical connections, linking multiple frontal cortical areas. *The Journal of Neuroscience* 22, 8117-8132 (2002).

13. D. Xiao, B. Zikopoulos, H. Barbas, Laminar and modular organization of prefrontal projections to multiple thalamic nuclei. *Neuroscience* 161, 1067-1081 (2009).
14. E. Y. Choi, S. L. Ding, S. N. Haber, Combinatorial inputs to the ventral striatum from the temporal cortex, frontal cortex, and amygdala: Implications for segmenting the striatum. *eNeuro* 4, ENEURO.0392-0317.2017 (2017).
15. H. F. Clarke, T. W. Robbins, A. C. Roberts, Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *The Journal of Neuroscience* 28, 10972-10982 (2008).
16. V. D. Costa, O. Dal Monte, D. R. Lucas, E. A. Murray, B. B. Averbeck, Amygdala and ventral striatum make distinct contributions to reinforcement learning. *Neuron* 92, 505-517 (2016).
17. C. A. Taswell, V. D. Costa, E. A. Murray, B. B. Averbeck, Ventral striatum's role in learning from gains and losses. *Proceedings of the National Academy of Sciences of the United States of America* 115, E12398-E12406 (2018).
18. A. S. Mitchell, P. G. F. Browning, M. G. Baxter, Neurotoxic lesions of the medial mediodorsal nucleus of the thalamus disrupt reinforcer devaluation effects in rhesus monkeys. *The Journal of Neuroscience* 27, 11289-11295 (2007).
19. S. Chakraborty, N. Kolling, M. E. Walton, A. S. Mitchell, Critical role for the mediodorsal thalamus in permitting rapid reward-guided updating in stochastic reward environments. *eLife* 5, e13588 (2016).
20. P. H. Rudebeck, R. C. Saunders, D. A. Lundgren, E. A. Murray, Specialized representations of value in the orbital and ventrolateral prefrontal cortex: desirability versus availability of outcomes. *Neuron* 95, 1208-1220.e1205 (2017).
21. B. N. Armbruster, X. Li, M. H. Pausch, S. Herlitze, B. L. Roth, Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proceedings of the National Academy of Sciences* 104, 5163-5168 (2007).
22. T. J. Stachniak, A. Ghosh, S. M. Sternson, Chemogenetic synaptic silencing of neural circuits localizes a hypothalamus→midbrain pathway for feeding behavior. *Neuron* 82, 797-808 (2014).
23. S. V. Mahler et al., Designer receptors show role for ventral pallidum input to ventral tegmental area in cocaine seeking. *Nature Neuroscience* 17, 577-585 (2014).
24. K. Oyama et al., Chemogenetic dissection of the primate prefronto-subcortical pathways for working memory and decision-making. *Science Advances* 7, eabg4246 (2021).
25. P. H. Rudebeck, R. C. Saunders, A. T. Prescott, L. S. Chau, E. A. Murray, Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nature Neuroscience* 16, 1140-1145 (2013).
26. M. J. Buckley et al., Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* 325, 52-58 (2009).
27. G. Schoenbaum, M. R. Roesch, T. A. Stalnaker, Y. K. Takahashi, A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Reviews Neuroscience* 10, 885-892 (2009).
28. M. M. Torregrossa, J. J. Quinn, J. R. Taylor, Impulsivity, compulsivity, and habit: The role of orbitofrontal cortex revisited. *Biological Psychiatry* 63, 253-255 (2008).
29. T. Setogawa et al., Neurons in the monkey orbitofrontal cortex mediate reward value computation and decision-making. *Communications Biology* 2, 126 (2019).
30. G. Schoenbaum, S. L. Nugent, M. P. Saddoris, B. Setlow, Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *NeuroReport* 13, 885-890 (2002).
31. R. Dias, T. W. Robbins, A. C. Roberts, Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* 380, 69-72 (1996).
32. P. H. Rudebeck, E. A. Murray, Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *The Journal of Neuroscience* 28, 8338-8343 (2008).
33. R. Rygula, S. C. Walker, H. F. Clarke, T. W. Robbins, A. C. Roberts, Differential contributions of the primate ventrolateral prefrontal and orbitofrontal cortex to serial reversal learning. *The Journal of Neuroscience* 30, 14552-14559 (2010).
34. P. H. Rudebeck, E. A. Murray, Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior. *The Journal of Neuroscience* 31, 10569-10578 (2011).
35. C. L. Pickens et al., Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *The Journal of Neuroscience* 23, 11078-11084 (2003).
36. E. A. Murray, E. J. Moylan, K. S. Saleem, B. M. Basile, J. Turchi, Specialized areas for value updating and goal selection in the primate orbitofrontal cortex. *eLife* 4, e11695 (2015).
37. D. Kim, G. Y. Park, O. D. JP, S. W. Lee, Task complexity interacts with state-space uncertainty in the arbitration between model-based and model-free learning. *Nature Communications* 10, 5738 (2019).
38. M. Malvaez, C. Shieh, M. D. Murphy, V. Y. Greenfield, K. M. Wassum, Distinct cortical-amygdala projections drive reward value encoding and retrieval. *Nature Neuroscience* 22, 762-769 (2019).
39. E. C. Fiuzat, S. E. V. Rhodes, E. A. Murray, The role of orbitofrontal-amygdala interactions in updating action-outcome valuations in macaques. *The Journal of Neuroscience* 37, 2463-2470 (2017).
40. S. M. Groman et al., Orbitofrontal circuits control multiple reinforcement-learning processes. *Neuron* 103, 734-746.e733 (2019).
41. A. Banerjee et al., Value-guided remapping of sensory circuits by lateral orbitofrontal cortex in reversal learning. *Nature* 585, 245-250 (2020).
42. G. E. Alexander, M. D. Crutcher, Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends in Neurosciences* 13, 266-271 (1990).
43. K. Kimura et al., A mosaic adeno-associated virus vector as a versatile tool that exhibits high levels of transgene expression and neuron specificity in primate brain. *Nature Communications* 14, 4762 (2023).
44. K. Oyama et al., Chemogenetic Disconnection between the Orbitofrontal Cortex and the Rostromedial Caudate Nucleus Disrupts Motivational Control of Goal-Directed Action. *The Journal of Neuroscience* 42, 6267-6275 (2022).
45. K. Oyama et al., Chemogenetic dissection of the primate prefronto-subcortical pathways for working memory and decision-making. *Science Advances* 7, eabg4246 (2021).
46. Y. Murata et al., Temporal Plasticity Involved in Recovery from Manual Dexterity Deficit after Motor Cortex Lesion in Macaque Monkeys. *The Journal of Neuroscience* 35, 84-95 (2015).
47. J.-A. Li et al., Quantum reinforcement learning during human decision-making. *Nature Human Behaviour* 4, 294-307 (2020).
48. K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, K. J. Friston, Bayesian model selection for group studies. *Neuroimage* 46, 1004-1017 (2009).

1



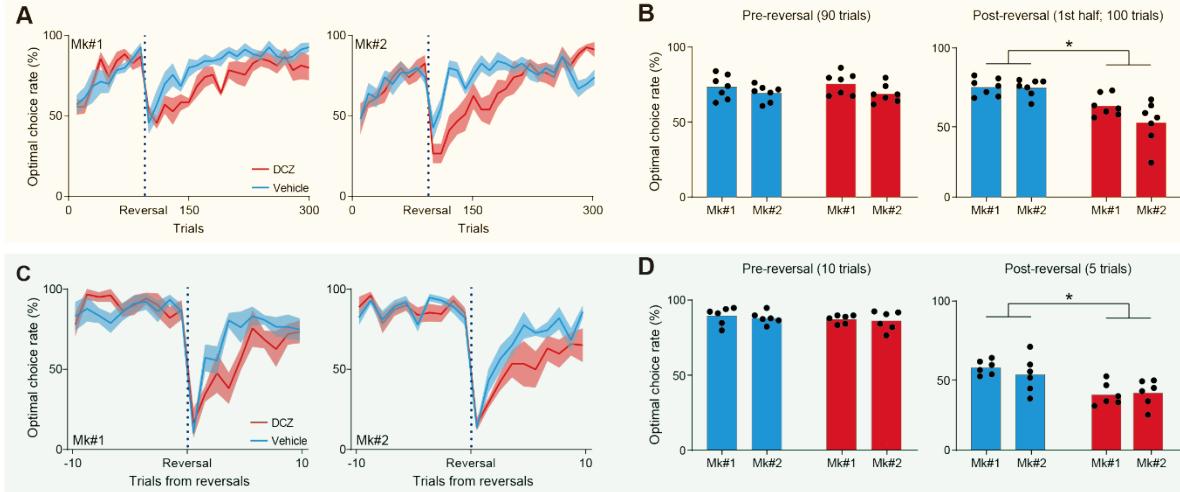
2

3 **Fig. S1 | Expression of hM4Di in the OFC and its terminal sites.**

4 (A to E) *In vivo* visualization of hM4Di expression in the OFC (A), rmCD (B), MDm (C),  
5 putamen (D), and amygdala (E) obtained from Mk#2. Images are coronal PET contrasts  
6 showing specific binding of [<sup>11</sup>C]DCZ (contrast: after the introduction of hM4Di minus  
7 before the introduction), overlaid by MR images from Mk#2. The middle row visualizes  
8 corresponding DAB-stained sections showing immunoreactivity against a reporter protein  
9 (AcGFP), and the bottom row shows an enlarged view of the areas marked with red  
10 rectangles. Scale bars: 5 mm.

11

12



13

### Fig. S2 | Effects of chemogenetic silencing of the bilateral OFC in each monkey.

14

(A) Behavioral performance for Mk#1 (left) and Mk#2 (right) on the NOVEL task ( $N = 7$  for each treatment). (B) Optimal choice rate for the pre-reversal phase (left, 90 trials; two-way ANOVA, treatment,  $F_{(1,24)} = 0.11, p = 0.74$ ; subject,  $F_{(1,24)} = 4.4, p = 4.6 \times 10^{-2}$ ; interaction,  $F_{(1,24)} = 0.21, p = 0.65$ ) and the 1st half of the post-reversal phase (right, 100 trials, treatment,  $F_{(1,24)} = 28.0, p = 2.0 \times 10^{-5}$ ; subject,  $F_{(1,24)} = 3.0, p = 0.10$ ; interaction,  $F_{(1,24)} = 2.5, p = 0.12$ ). (C) Behavioral performance for Mk#1 (left) and Mk#2 (right) on the FAMILIAR task ( $N = 6$  for each treatment). (D) Optimal choice rate for the pre-reversal phase (left, 10 trials; two-way ANOVA, treatment,  $F_{(1,20)} = 0.36, p = 0.56$ ; subject,  $F_{(1,20)} = 0.9, p = 0.34$ ; interaction,  $F_{(1,20)} = 0.009, p = 0.93$ ) and the post-reversal phase (right, 5 trials, treatment,  $F_{(1,20)} = 17.8, p = 4.2 \times 10^{-4}$ ; subject,  $F_{(1,20)} = 0.25, p = 0.62$ ; interaction,  $F_{(1,20)} = 0.68, p = 0.42$ ). Asterisks:  $p < 0.05$  for significant main effect of treatment.

15

16

17

18

19

20

21

22

23

24

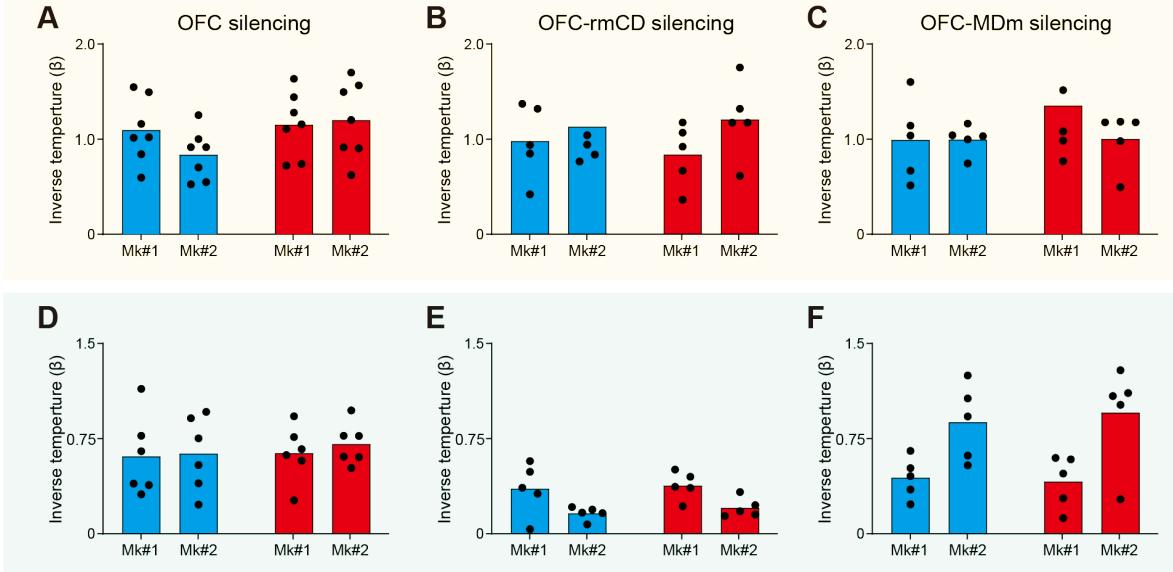
25

26

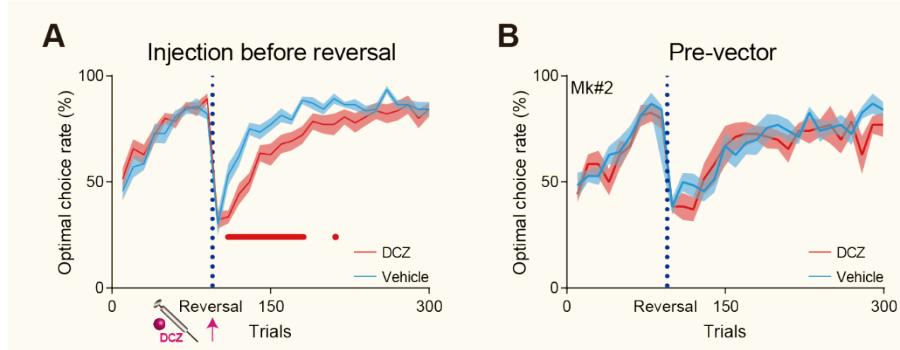
27

28 **Fig. S3 | Estimated inverse temperature by model fitting analysis.**

29 (A to C) Inverse temperatures estimated by fitting performance on the NOVEL task to the  
30 EXP model after silencing the OFC (A) (two-way ANOVA, treatment,  $F_{(1,24)} = 2.7, p =$   
31 0.11; subject,  $F_{(1,24)} = 0.69, p = 0.41$ ; interaction,  $F_{(1,24)} = 1.4, p = 0.24$ ), the OFC-rmCD  
32 pathway (B) (treatment,  $F_{(1,16)} = 0.03, p = 0.86$ ; subject,  $F_{(1,16)} = 1.9, p = 0.18$ ; interaction,  
33  $F_{(1,16)} = 0.33, p = 0.58$ ), or the OFC-MDm pathway (C) (treatment,  $F_{(1,16)} = 0.94, p = 0.35$ ;  
34 subject,  $F_{(1,16)} = 0.85, p = 0.37$ ; interaction,  $F_{(1,16)} = 0.87, p = 0.37$ ) for control vehicle  
35 (cyan) and DCZ treatment (red) in each monkey. Only data for the post-reversal phase are  
36 shown. (D to F) Inverse temperatures estimated by fitting to the performance on the  
37 FAMILIAR task to the INF model after silencing of the OFC (D) (two-way ANOVA,  
38 treatment,  $F_{(1,20)} = 0.21, p = 0.65$ ; subject,  $F_{(1,20)} = 0.24, p = 0.63$ ; interaction,  $F_{(1,20)} =$   
39 0.054,  $p = 0.82$ ), the OFC-rmCD pathway (E) (treatment,  $F_{(1,16)} = 0.39, p = 0.54$ ; subject,  
40  $F_{(1,16)} = 11.0, p = 4.5 \times 10^{-3}$ ; interaction,  $F_{(1,16)} = 0.026, p = 0.87$ ), or the OFC-MDm  
41 pathway (F) (treatment,  $F_{(1,16)} = 0.034, p = 0.86$ ; subject,  $F_{(1,16)} = 15.4, p = 1.2 \times 10^{-3}$ ;  
42 interaction,  $F_{(1,16)} = 0.17, p = 0.68$ ).  
43  
44  
45



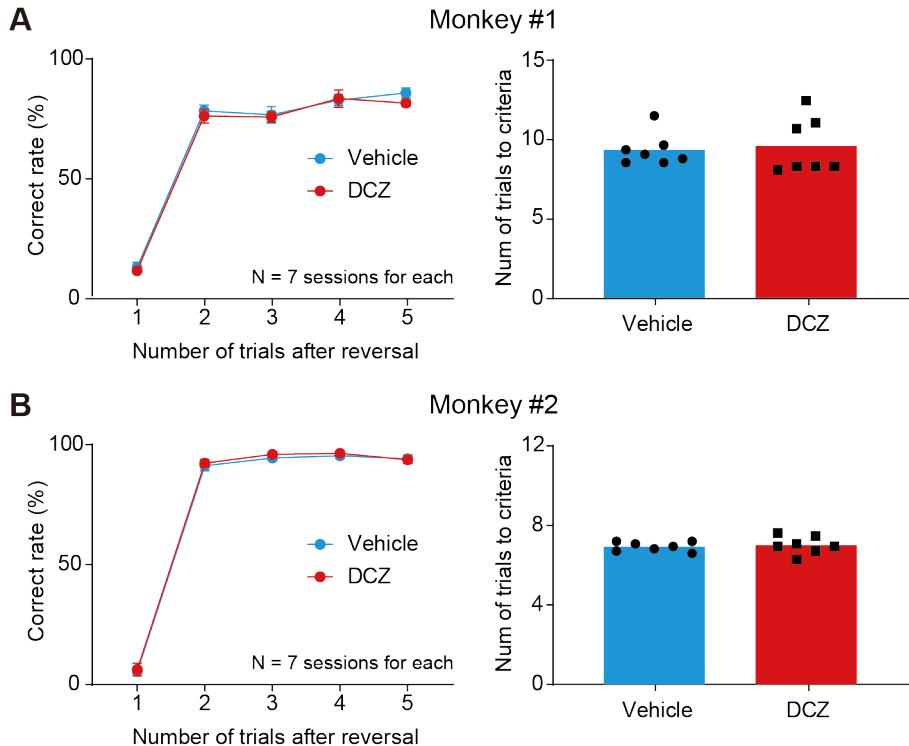
46



47 **Fig. S4 | Effects of intramuscular DCZ administration just before the reversal and**  
48 **without hM4Di expression on the NOVEL task performance.**

49 (A) Behavioral performance on the NOVEL task (N = 7 for each treatment in each  
50 monkey) when DCZ was administered intramuscularly just before reversal. The task was  
51 interrupted for 5 min and DCZ was administered at the start of the 5-min break. (B)  
52 Behavioral performance on the NOVEL task (N = 7 for each treatment, Mk#2) before the  
53 introduction of hM4Di.

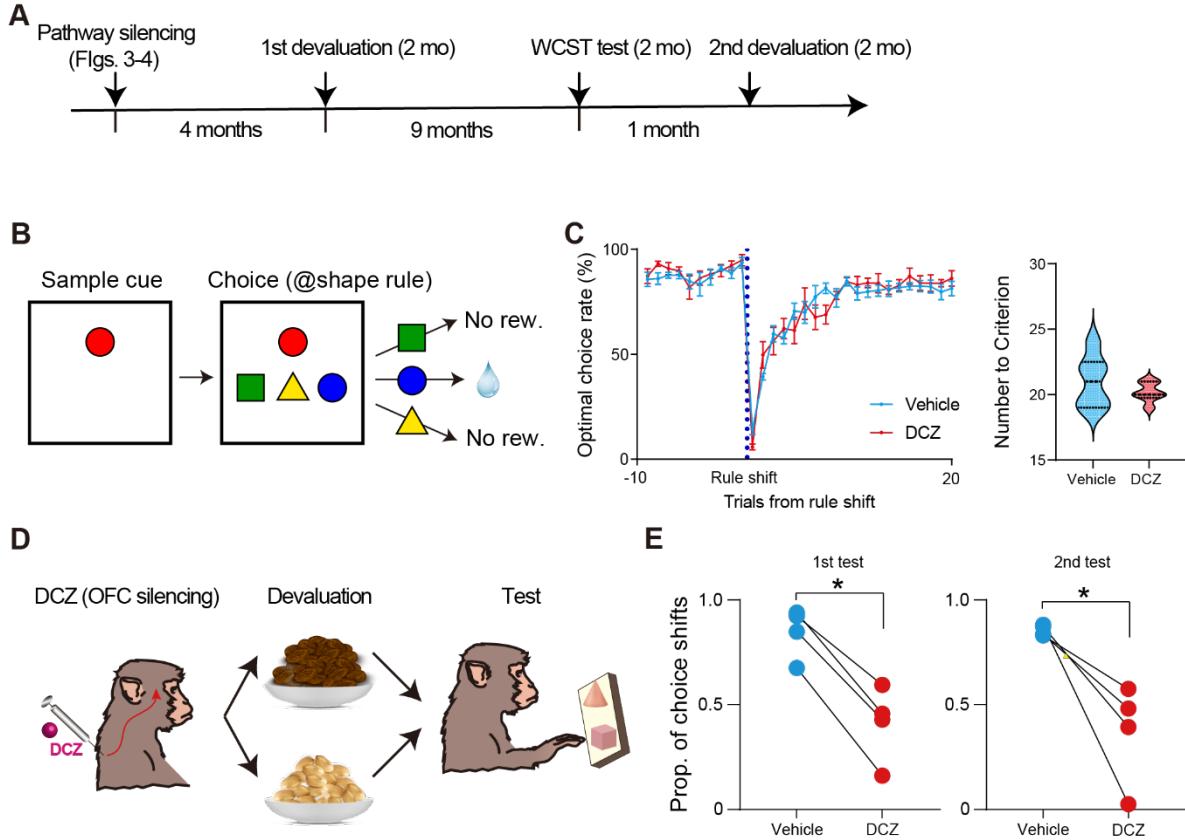
54



55

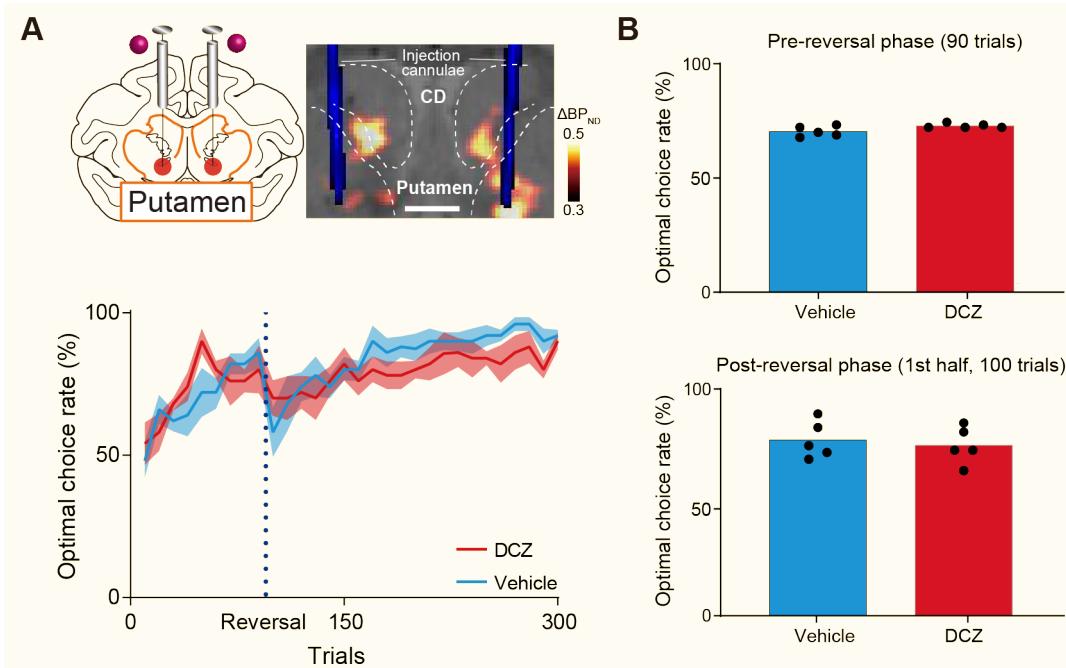
56 **Fig. S5 | Effect of chemogenetically silencing the bilateral OFC on the two-arm reversal**  
57 **learning task.**

58 **(A and B)** Correct rate as a function of the number of trials after reversal (left) and the  
59 number of trials to reach criteria (right) for Mk#1 **(A)** and Mk#2 **(B)**. There was no  
60 significant difference between vehicle and DCZ injection (Two-tailed Welch's t-test,  
61 Mk#1,  $t_{9.7} = 0.33, p = 0.75$ ; Mk#2,  $t_{9.1} = 0.37, p = 0.72$ ).  
62



63  
64 **Fig. S6 | Effect of chemogenetically silencing the bilateral OFC on performance for the**  
65 **Wisconsin Card Sorting Test and the devaluation test.**

66 Schema of the Wisconsin Card Sorting Test. (C) Optimal choice rate ( $N = 6$  for each  
67 treatment) as a function of trials for rule shift (left) and the number of trials to reach the  
68 criterion for rule shift (right). There was no significant difference in the number needed to  
69 reach criterion between vehicle and DCZ injection conditions (Two-tailed Welch's t-test,  
70  $t_{6.4} = 0.96, p = 0.37$ ). The horizontal lines in each violin plot show the quartiles of the  
71 distributions. (D) Schema for the devaluation test. (E) Performance on the devaluation test  
72 for the 1st (left) and 2nd schedule (right), respectively ( $N = 4$  for each treatment for both  
73 schedules). There was a significant difference in performance between vehicle and DCZ  
74 injections for both schedules (Two-tailed Welch's t-test, 1st,  $t_{3.1} = 4.0, p = 2.7 \times 10^{-2}$ ).  
75



76

77 **Fig. S7 | Effects of chemogenetically silencing the OFC-mPut pathway on NOVEL**  
78 **task performance.**

79 (A) Chemogenetically silencing the OFC-mPut pathway by local DCZ infusion into  
80 bilateral mPut, specifically at hM4Di-positive OFC terminals sites (top), and behavioral  
81 performance (bottom). N = 5 sessions for each treatment. Conventions are the same as in  
82 Fig. 2. (B) Averaged optimal choice rate in different task phases (Two-tailed Welch's t-test,  
83 pre-reversal phase:  $t_{5.4} = 2.2, p = 0.08$ ; 1st half of post-reversal phase:  $t_{8.0} = 0.46, p = 0.66$ ).  
84  
85

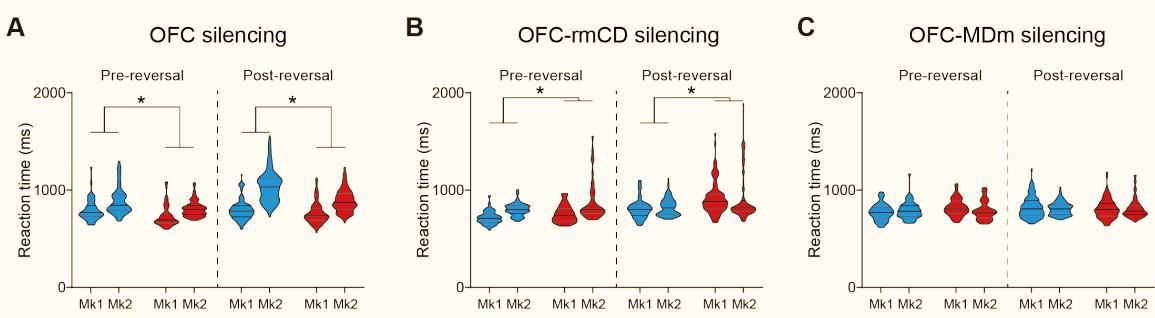
86

87 **Fig. S8 | Effects of chemogenetically silencing the OFC, OFC-rmCD pathway, and**  
88 **OFC-MDm pathway on reaction time during the NOVEL task.**

89 (A to C) Reaction time in the pre-reversal phase (left) and post-reversal phase (right) after  
90 silencing the OFC (A), OFC-rmCD pathway (B), and OFC-MDm pathway (C) with DCZ  
91 treatment (red) or after control vehicle (cyan, no silencing) in each monkey. In both task  
92 phases, reaction times increased significantly after OFC silencing (pre-reversal phase:  
93 treatment,  $F_{(1,248)} = 25.3, p = 9.5 \times 10^{-7}$ ; subject,  $F_{(1,248)} = 38.0, p = 2.8 \times 10^{-9}$ ; interaction,  
94  $F_{(1,248)} = 0.62, p = 0.43$ ; post-reversal phase: treatment,  $F_{(1,584)} = 74.1, p = 6.8 \times 10^{-17}$ ;  
95 subject,  $F_{(1,584)} = 325.7, p = 3.4 \times 10^{-58}$ ; interaction,  $F_{(1,584)} = 74.1, p = 7.3 \times 10^{-10}$ ), decrease  
96 significantly after silencing the OFC-rmCD pathway (pre-reversal phase: treatment,  $F_{(1,176)} = 10.5, p = 1.5 \times 10^{-3}$ ; subject,  $F_{(1,176)} = 32.6, p = 4.7 \times 10^{-8}$ ; interaction,  $F_{(1,176)} = 0.77, p = 0.38$ ; post-reversal phase: treatment,  $F_{(1,416)} = 37.8, p = 1.9 \times 10^{-9}$ ; subject,  $F_{(1,416)} = 0.34, p = 0.56$ ; interaction,  $F_{(1,416)} = 0.39, p = 0.53$ ), and did not change after silencing the OFC-  
100 MDm pathway (pre-reversal phase: treatment,  $F_{(1,176)} = 2.7, p = 0.10$ ; subject,  $F_{(1,176)} = 0.14, p = 0.71$ ; interaction,  $F_{(1,176)} = 3.7, p = 0.06$ ; post-reversal phase: treatment,  $F_{(1,416)} = 0.76, p = 0.38$ ; subject,  $F_{(1,416)} = 0.22, p = 0.64$ ; interaction,  $F_{(1,416)} = 0.58, p = 0.45$ ). The  
101 horizontal lines in each violin plot show the quartiles of the distributions. Asterisks:  $p <$   
102  $0.05$  for significant main effect of treatment.  
103

104

105



106  
107

108 **Fig. S9 | The effects of silencing the OFC, OFC-rmCD pathway, and OFC-MDm**  
109 **pathway on the sensitivity to past outcomes during the FAMILIAR task.**

110 (A to C) Averaged optimal choice rates for trials after the reversal in the FAMILIAR task  
111 following positive (left) and negative (right) outcomes after OFC silencing (A) (positive:  
112 treatment,  $F_{(1,20)} = 3.7$ ,  $p = 0.069$ ; subject,  $F_{(1,20)} = 1.1$ ,  $p = 0.31$ ; interaction,  $F_{(1,20)} = 0.47$ ,  $p$   
113 = 0.50; negative: treatment,  $F_{(1,20)} = 7.2$ ,  $p = 1.4 \times 10^{-2}$ ; subject,  $F_{(1,20)} = 0.33$ ,  $p = 0.57$ ;  
114 interaction,  $F_{(1,20)} = 0.36$ ,  $p = 0.55$ ), OFC-rmCD silencing (B) (positive: treatment,  $F_{(1,16)} =$   
115 1.0,  $p = 0.32$ ; subject,  $F_{(1,16)} = 6.6$ ,  $p = 2.1 \times 10^{-2}$ ; interaction,  $F_{(1,16)} = 0.58$ ,  $p = 0.46$ ;  
116 negative: treatment,  $F_{(1,16)} = 0.21$ ,  $p = 0.65$ ; subject,  $F_{(1,16)} = 15.1$ ,  $1.3 \times 10^{-3}$ ; interaction,  
117  $F_{(1,16)} = 0.0004$ ,  $p = 0.99$ ), and OFC-MDm silencing (C) (positive: treatment,  $F_{(1,16)} = 0.40$ ,  
118  $p = 0.54$ ; subject,  $F_{(1,16)} = 2.6$ ,  $p = 0.13$ ; interaction,  $F_{(1,16)} = 0.10$ ,  $p = 0.76$ ; negative:  
119 treatment,  $F_{(1,16)} = 6.0$ ,  $p = 2.6 \times 10^{-2}$ ; subject,  $F_{(1,16)} = 18.0$ ,  $p = 6.0 \times 10^{-4}$ ; interaction,  
120  $F_{(1,16)} = 0.17$ ,  $p = 0.68$ ).  
121

