Region- and stage-specific cortical plasticity shapes the multifaceted process of habit formation

Nozomi Asaoka* and Yasunori Hayashi*

Department of Pharmacology, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan

Running title: **Cortical plasticity and habit formation**

* Co-corresponding authors:
Nozomi Asaoka, PhD
Email: asaoka.nozomi.3s@kyoto-u.ac.jp

Yasunori Hayashi, MD, PhD
E-mail: yhayashi-tky@umin.ac.jp

ORCID
0000-0003-4854-2100 (NA)
0000-0002-7560-3004 (YH)

Summary

Habits are crucial for shaping our normal and pathological behaviors. They emerge through a transition in decision-making strategy, from flexible adaptation to current demands (goal-directed behavior) to automatic repeats of previous behaviors (habitual behavior). This transition of decision-making strategy induces changes in an underlying cognitive process for selecting an action while superficially maintaining the execution of the selected action. However, the neurocircuit mechanisms supporting these changes remain unexplored. By developing a behavioral paradigm capable of inducing this transition within a defined time window, we performed a longitudinal analysis of the neuronal circuits involved in this process using *ex vivo* electrophysiology, optogenetic erasure of synaptic plasticity and *in vivo* Ca^{2+}-imaging. We found that the decision-making strategy and action execution are orthogonally controlled by plasticity in distinct cortical pathways. Erasure of these changes independently alter the decision-making strategy or the action execution without influencing each other, supporting a dichotomous circuit mechanism. This study underscores the critical role of pathway-specific cortical plasticity in regulating the multifaceted process shaping habitual behavior, offering valuable insights into the pathological mechanisms of maladaptive habits, such as compulsive disorders.

Keywords

Habitual behavior, decision making, synaptic plasticity, cortex

Introduction

The decision on a behavior for a given situation represents a critical aspect of cognitive function. Two strategies have been identified as central to this decision-making: the goal-directed and habitual strategies[1–5]. Generally, a behavior is learnt with the goal-directed strategy, where a subject flexibly makes a decision on what to do, according to the expected outcome of the behavior as well as internal needs[1,2]. As the behavior is repeated in a particular context, it transits into a habit, where the habitual strategy dominates over the goal-directed strategy[3,4]. Under the habitual strategy, subjects select a behavior in response to a specific context without taking the value of outcome into account[5], thereby allowing the behavior to be stably repeated despite motivational changes. Owing to this, the habitual strategy is effective to maintain beneficial but effortful behaviors over time. At the same time, it can trigger uncontrolled repetition of behavior under pathological conditions such as obsessive-compulsive disorder and addiction[2,6]. However, the underlying neural mechanisms that govern the transition from the goal-directed strategy to the habitual strategy, as observed in real-life habit formation, remain to be elucidated.

The goal-directed and habitual strategies are supported by distinct forms of associative learning, termed action-outcome and stimulus-response associations, respectively[2,3,7]. The predominance of goal-directed versus habitual strategies depends on the relative strength of their respective associative learning[2,7]. It has been proposed that repetition of a behavior leads to habit formation through strengthening of the stimulus-response association, while concomitantly weakening the subject's attention to the action-outcome association[2,3,7,8]. In other words, when a behavior is initially acquired as goal-directed one, the neural mechanisms that diminish the reliance on the action-outcome association may be also critical for habit formation. In light of these concepts, our research focused on the dynamics of the goal-directed strategy from the early stages of behavior learning to habit formation.

From the perspective of behavior acquisition as a reinforcement learning, the probability of engaging

in a specific behavior is increased as learning progresses[3,5]. In this framework, behavior is evaluated not only by the qualitative criterion of whether goal-directed or habitual strategies are predominant, but also by quantitative analysis of execution, such as frequency and duration of learned action[9]. However, control mechanisms of execution levels under the habitual strategy remain elusive. For ensuring the efficacy and benefits of a habit, maintenance of a consistent level of execution is also crucial. This poses a challenge to the subjects. When the goal-directed strategy is dominant, execution is optimized to maximize the cost-benefit ratio, which is highly dependent on motivational level[9]. To sustain its efficacy and ensure the continued benefit of a habit even when motivation (or its impact on decision strategy) wanes, the same level of execution must be preserved before and after the transitioning to the habitual strategy. These two aspects of habit, decision-making on action selection and following behavior execution are conceptualized as distinct computational steps, that are independently controlled in human psychological studies[10,11]. This motivated us to dissect out the neuronal mechanisms responsible for these two aspects of behavior control.

Previous studies have demonstrated different brain regions support the goal-directed and habitual strategies[5,12,13]. The prefrontal cortex and its downstream subcortical regions are implicated in goal-directed strategy. When subjects predominantly rely on the habitual strategy, the influence of the sensorimotor cortex and its downstream targets is prominent instead of the prefrontal cortex pathway[5,12,14]. These distinct activation patterns by the dominance of the goal-directed or habit strategies have led us to postulate that neuronal plasticity underlies the transition of the decision-making strategy as well as maintaining the execution level.

To address this, we developed a novel behavior task, which allows us to dissect the neuronal mechanisms responsible for the transitions in decision-making strategies together with determining the behavior execution level in the same cohort of animals. We combined this task with *ex vivo* slice recordings, *in vivo* Ca^{2+} imaging and a recently-developed optogenetic method to erase long-term potentiation (LTP). Our analyses revealed that predominance of the goal-directed strategy and the

execution level are orthogonally represented as potentiation of synaptic transmission in different prefrontal cortical regions. By targeting region specific plastic changes, we were able to change the decision-making strategy without affecting the executive level or *vice versa*, supporting the notion that the decision on a behavior and the execution of the behavior are distinct processes that involve different circuits. Our findings not only extend the current understanding of the neuronal mechanisms of habit formation, but also indicate that targeting cortical plasticity could offer a therapeutic strategy for addressing maladaptive habits, such as obsessive-compulsive disorder and addiction.

Results

Experimental induction of a transition between goal-directed and habitual behaviors

Previously studies have demonstrated that training protocols with ratio-based and interval-based operant conditioning tasks have contrasting outcomes on animal behaviors[3,5,15]. With ratio-based tasks, where the rewards are delivered after a certain number of lever presses, animals acquire goal-directed lever pressing behavior. In contrast, with interval-based tasks, where rewards are delivered on the first lever press after a specific amount of time has elapsed since the prior reward, animals develop habitual lever pressing behavior[16–18]. The interval-based tasks are designed to allow the number and frequency of lever pressing to be less contingent on reward deliveries than the ratio-based tasks. In other words, the timing of reward deliveries is less affected by subjects' efforts on lever pressing in the interval-based tasks. This task design, along with habit formation through over-repetition, enables subjects to predict reward delivery without computation of the action-outcome association; therefore, resulting in the predominance of the habitual (stimulus-response) strategy over the goal-directed strategy[3,19].

To date, studies have used these two protocols separately, to monitor the acquisition of a behavior controlled by either goal-directed or habitual strategies [3,5]. However, these separate approaches make identifying the timing of the transition from goal-directed to habitual behaviors extremely difficult, and thus don't allow for the study of the neural mechanisms underlying this transition. To overcome this, we performed both ratio- and interval-based tasks in a sequential fashion in the same cohort of mice. First, goal-directed behavior was established with a ratio-based task, after which habitual behavior was invoked by switching to an interval-based task, allowing us to track within-subject transitions of decision-making strategy and action execution.

Animals were first trained with the continuous reinforcement (CRF) task in an operant chamber for 3 days (days 1–3), where each lever press was rewarded with a drop of sucrose solution from the reward port (Fig. 1A). Following this, they were trained with a ratio-based, variable ratio (VR) task

for 6 days (days 4–9), being rewarded after 10 presses on average on days 4–5, and 20 presses on days 6–9. The lever press rate during VR training increased, indicating successful learning of the action-outcome (lever-reward) association[3,5] (Fig. 1B). On day 10, the task was seamlessly switched to an interval-based, variable interval (VI) task, where animals were rewarded on the first lever press after a fluctuating time interval from the last reward (ranging from 30 to 90 s, average 60 s). When switched to the VI task, mice slightly but statistically significantly declined the frequency of lever pressing, compared with animals that continued the VR task (VR+VR group) (Fig. 1B).

To confirm the task design allowed for a transition between the goal-directed to habitual strategy, we conducted two tests: a devaluation test and an omission test[20–22]. In the devaluation test, mice were given free access to either a sucrose solution (devalued condition), or regular rodent chow (valued condition) for 30 minutes, before being subjected to the lever-press task (Fig. 1C). On the following day, the same mice were tested in the opposite condition. When the goal-directed strategy is predominant, mice are expected to reduce the number of lever presses in the devalued condition compared to the valued condition, as they are already satiated with the sucrose solution. Conversely, under the habitual strategy, the number of lever presses should be comparable between conditions, because the reward value no longer discourages the decision. We defined the devaluation index for each mouse as the difference in lever pressing between the two conditions normalized by the total number of lever presses (see methods), which results in positive when the subjects rely on the goal-directed strategy and approaches zero when they employ the habitual strategy[23]. When the devaluation test was conducted after VR training, there was a significant reduction in lever pressing in the devalued condition resulting in a positive devaluation index (Fig. 1D–F), confirming that mice were employing a goal-directed strategy[5,22]. However, when mice were trained with the VI task for an additional 4 days, they executed a comparable number of lever presses in both the valued and devalued conditions, leading to devaluation indices near zero, indicating they had changed their strategy to a habitual one. This was not simply the result of prolonged training, because mice that

underwent 13-days of VR training (VR+VR) still exhibited a reduction in the number of lever presses in the devalued condition (Fig. 1D & F). In contrast to lever pressing, significant decreases in the numbers of nose pokes to the reward delivery port were observed in the devalued condition for all groups, suggesting that their motivation towards sucrose rewards was reduced irrespective of the training conditions (Fig. S1A & B). Therefore, it can be inferred that the lack of inhibitory effects of devaluation on lever pressing in the VR+VI mice was not due to impaired satiety-mediated outcome devaluation, but rather due to their decision strategy, which did not take into account the current value of sucrose rewards. The effects of devaluation in the VR+VI mice were comparable to those of mice trained with 13-days of VI tasks (Fig. S1C–I). Consequently, these results confirm that the VR→VI switching, rather than simply extended training, leads to the formation of habitual behavior.

Habit formation was also assessed in the omission test[22]. Following the completion of VR+VI training, sucrose solution was delivered automatically every 20 seconds when the lever was not pressed, while lever pressing triggered a delay in reward delivery, thereby degrading the association between lever pressing and sucrose delivery (Fig. S2A). Mice utilising a goal-directed strategy are expected to diminish the number of lever presses more quickly than those adopting a habitual strategy [22]. Consistent with this idea, the number of lever presses decreased slower in the VR+VI group, compared with the VR+VR group (Fig. S2B & C).

Taken together, these data indicate that implementing a VI training period, following VR training (VR+VI task) converts an established goal-directed behavior into habitual behavior, thereby providing a versatile paradigm to allow for the detailed analysis of the circuit mechanisms underlying the transition of the predominant decision strategy from goal-directed to habitual one.

Execution of habitual behavior is controlled independently from of the bias toward habitual decision-making

Our VR+VI task data revealed there is considerable variation in within-subject changes in the lever pressing behavior associated with the transition from the goal-directed to habitual decision strategy (Fig. 1G–O). We hypothesized that this variation may prove valuable for identify the neuronal mechanisms that determine execution level. An execution index for each mouse was calculated by normalizing the difference in the number of lever presses between days 9 (final VR session) and 13 (final VI session), by the total number of lever presses in the two sessions (Fig. 1G, methods). The execution index varied among mice trained in the VR+VI task from positive to negative with a median of -0.16 (max: 0.44, min: -0.58, standard deviation (SD): 0.21). In contrast, mice trained with the VR+VR protocol showed significantly higher execution indices with a median of 0.11 (max: 0.71, min: -0.09, SD: 0.18), indicating that the majority of mice maintained or increased their lever pressing rate after 4-days of additional VR training (Fig. 1G).

Next, for further analysis, the VR+VI group was subdivided into two groups; VR+VI_{High} and VR+VI_{Low}, based on execution indices falling either side of the median value (-0.16) (Fig. 1G & H). The lever press rate on day 9 was comparable between the VR+VI_{High} and VR+VI_{Low} groups and did not affect this classification (Fig. 1I & J). Likewise, the number of lever press bouts (see methods for definition) was comparable on day 9 between the VR+VI_{High} and VR+VI_{Low} groups, however on day 13 it was significantly decreased in the VR+VI_{Low} group (Fig. 1K). These data indicate there was no significant difference in behavior between the VR+VI_{High} and VR+VI_{Low} groups during goal-directed behavior, prior to the VR→VI switching. Similarly, there was no significant change in the number of nose pokes on both days 9 and 13 (Fig. 1L), suggesting reward value was comparable despite the VR→VI switching protocol. These observations indicate that differences in execution index resulted from habit formation, rather than reflecting pre-existing variability.

To test if the strength of habitual strategy reflects the variation observed in execution indices, we compared the results of the devaluation and omission tests between the VR+VI_{High} and VR+VI_{Low} groups. In the devaluation test, both groups displayed a significant inhibitory effect before VR→VI

switching, but not after VR→VI switching (Fig. 1M & N). Both before and after switching, there was no significant correlation between the devaluation index and execution index (Fig. 1O), suggesting that the strength of goal-directed or habitual strategy had no impact on the maintenance of behavioral execution levels. In the omission test, the decay in lever pressing was comparable across the VR+VI_{High} and VR+VI_{Low} groups, both of which exhibited a slower decay than the VR+VR group, consistent with the idea that the strength of the habitual strategy does not correlate with execution level (Fig. S2D–H).

We next tested if the individual differences in execution index arise from other behavioral factors by using a T-maze spatial discrimination and reversal learning test and marble burying test, which can measure learning ability, cognitive flexibility and compulsive-like behavior, respectively, after completion of the VR+VI task (Fig. S3). In all experiments, there was no significant difference between the VR+VI_{High} and VR+VI_{Low} groups, suggesting that the differences in execution indices were determined by the daily training, rather than the individual's general characteristics of cognition or decision-making.

Collectively, these results suggest that the VR→VI switching protocol induces a transition from the goal-directed to habitual strategy, thought the bias toward habitual strategy was not correlated with behavioral execution, and thus is likely determined by an independent mechanism.

Cortical region-specific potentiation of synaptic responses in specific task stages

Since repeated training was required for both behavioral learning and habit formation, we hypothesized that training induce plastic changes in synaptic transmission, which may correlate the dynamics of goal-directed/habitual strategy predominance and behavior execution. To test this, excitatory postsynaptic currents (EPSC) were recorded *ex vivo* from pyramidal neurons in the prefrontal cortical regions; anterior cingulate, lateral and medial orbitofrontal cortices (ACC, LOFC and MOFC), which have been implicated in reinforcement learning and decision-making[12,24]. Mice

were sacrificed after CRF (day 3), VR (day 9) and VR+VR or VR+VI (day 13) training, after which AMPA/NMDA ratios were recorded from pyramidal neurons in layer 2/3 (L2/3) and layer 5 (L5), as an index of potentiation of AMPA receptor-mediated excitatory synaptic transmission, whose increase indicates recent induction of synaptic potentiation[25,26] (Fig. 2A-F).

Following CRF training, there was a specific increase in AMPA/NMDA ratio in MOFC L2/3 neurons, but not in other regions or layers (Fig. 2A–F). However, in the group that underwent subsequent VR training, this increase was diminished. Instead, there were increases in the AMPA/NMDA ratios in both L5 neurons of the ACC and LOFC (Fig. 2B, D). This suggests that as goal-directed behavior is progressively acquired, synaptic potentiation begins in MOFC, but transitions to encompass primarily the ACC and LOFC.

Following VR→VI switching (Fig. 2G–L), the AMPA/NMDA ratio in ACC L5 neurons returned to control levels in both the VR+VI_{High} and VR+VI_{Low} groups, whereas in the VR+VR group it remained elevated (Fig. 2G, H). However, the VR+VI_{High} group showed significantly higher AMPA/NMDA values in the LOFC L5 neurons, but not L2, compared with both the VR+VR and VR+VI_{Low} groups (Fig. 2I & J). In the MOFC, there were no significant differences in AMPA/NMDA ratios among groups (Fig. 2K & L). Finally, when comparing individual animals, we observed a significant correlation between the AMPA/NMDA ratio of LOFC L5 neurons and execution index (Fig. 2M-R).

To confirm the findings observed with evoked EPSCs, we recorded Sr^{2+}-induced AMPAR-mediated asynchronous EPSCs (aEPSCs) (Fig. S4). The results were consistent with the evoked EPSC data, except for a high aEPSC amplitude observed from LOFC L5 neurons in the VR+VR group (Fig. S4F), which contradict the evoked EPSCs we observed which were comparable to control levels. In the late stage of LTP, NMDA currents are potentiated following potentiation of AMPA current, resulting reset of the AMPA/NMDA ratio[26]. Therefore, it is possible that NMDAR inputs to those LOFC L5 neurons had also been potentiated by prolonged VR+VR training, potentially leading to an overall decrease in AMPA/NMDA ratios, compared with the VR group.

We also tested for changes in presynaptic release probability using the paired-pulse ratio (PPR; Fig. S5). There were no significant differences in the PPR in all regions tested, except for a significant decrease in the ACC L5 pyramidal neurons of VR groups, suggesting that presynaptic release probability is also increased.

Together, our observations suggest that region-specific plastic changes in the cortex potentially underlie transitions of the goal-directed and habitual strategy, as well as behavior execution levels. Acquisition of a goal-directed behavior potentiated the AMPAR-mediated synaptic transmission in the ACC L5 neurons, which was depotentiated by habit formation. Instead, in the VR+VI group, task-induced potentiation of the excitatory synaptic transmission in the LOFC L5 neurons was consecutively induced after habit formation, in a manner correlating with the execution index.

Chemogenetic inhibition of ACC facilitates habit formation while excitation reverts habitual to goal-directed behavior

To test if the increased excitability of ACC neurons is required for the predominance of the goal-directed over habitual strategy, we employed a chemogenetic approach. The inhibitory receptor hM4Di was expressed in ACC and mice were subsequently trained in the VR task for acquisition of a goal-directed behavior. Next, mice were administered with either vehicle or clozapine-*N*-oxide (CNO) to inhibit ACC neurons prior to the devaluation test (Fig. 3A–D). After the first devaluation test, mice were re-trained with the VR protocol for a further day, then underwent a second devaluation test with the reverse drug condition. Lever pressing in mice that received the vehicle was significantly decreased in the devaluation test, however the CNO treated group showed no significant changes between the valued and devalued conditions, indicative of a habitual strategy predominance. Conversely, when the activity of ACC neurons was increased with hM3Dq after the transition to habitual behavior by the VR+VI task, devaluation induced a significant reduction in lever pressing in the devaluated condition, indicating that goal-directed strategy could be reverted to predominant via

artificial ACC modulation (Fig. 3E–H). Additionally, similar effects of the ACC activation on the devaluation test were consistently observed in mice trained only with the VI task for 6 days, suggesting the contribution of ACC on the goal-directed strategy regardless of the training paradigms (Fig. S6A–C).

Interestingly, although VR training induced synaptic potentiation in the LOFC (Fig. 2J, P), chemogenetic inhibition did not affect the devaluation-induced response inhibition, unlike ACC inhibition (Fig. 3I–L, Fig. S6D). Similarly, chemogenetic manipulation of LOFC during the omission test didn't significantly alter lever pressing when comparing the vehicle- and CNO-treated hM4Di or hM3Dq groups (Fig. S7A-D). Moreover, in both conditions, there was no significant correlation between execution index and reduction of lever press rate in the omission test (Fig. S7E–G).

In all chemogenetic manipulations, CNO administration did not affect preference and satiety for sucrose rewards in the devaluation processes, because no significant changes were detected in sucrose/chow intake balance and reduction in nose poking towards the reward delivery port in the devalued condition (Fig. S6D–H).

Together, these observations indicate that the suppression of ACC activity can convert behavior from a goal-directed to habitual strategy, whereas activation drives goal-directed behavior. By contrast, the excitability of LOFC does not lead to shifts in decision strategy.

Projection-specific plasticity of ACC neurons in specific training stages

ACC neurons are a heterogenous population in terms of both gene expression and projection patterns[27,28]. To determine whether training-induced synaptic potentiation occurs in a specific subclass of ACC neurons, which target a particular region, or if it is more global change encompassing the entire ACC, we injected retrograde AAV encoding tdTomato into the major targets: the central striatum (CS), retrosplenial cortex (RSC), and the basolateral amygdala (BLA) to label of each subclass of ACC projection neurons[29–33] (Fig. 4A). Following this, training-induced synaptic

potentiation was assessed in tdTomato-positive L5 pyramidal neurons by recording the AMPA/NMDA ratio (Fig. 4B–J). A significant increase in the AMPA/NMDA ratio after VR training was detected only in the RSC-projecting L5 neurons (Fig. 4C). After VR→VI switching, the AMPA/NMDA ratio in this class of neurons returned to control levels, in both the VR+VI_{High} and VR+VI_{Low} groups, whereas the VR+VR group remained elevated (Fig. 4F). There was no significant correlation between the AMPA/NMDA ratio and execution index in each subclass (Fig. 4H–J). These results indicate that synaptic potentiation in ACC RSC-projection neurons plays an essential role in the acquisition of goal-directed behavior and that its de-potentiation is required for habit formation.

Optical erasure of synaptic potentiation in RSC-projecting ACC neurons facilitates the transition to habit

To further test this idea, we employed the recently-developed optogenetic tool, cofilin-SuperNova (CFL-SN) to erase long-term potentiation (LTP). Cofilin is an actin-binding protein that accumulates in dendritic spines and stabilizes filamentous actin after the induction of LTP[34]. Inactivation of cofilin by chromophore-assisted light inactivation (CALI) using a photosensitizer protein SuperNova (SN) erases LTP *post hoc,* without altering baseline transmission[35]. CFL-SN was expressed in the RSC-projecting ACC neurons to determine if *post hoc* LTP erasure would shift the predominant decision strategy to habitual from a goal-directed one. To this end, retrograde AAV was used to express CFL-SN, or SN as a control in the RSC-projecting ACC neurons bilaterally (Fig. 4K & L). After each VR session (days 4–9), light was delivered to the ACC through optical fibers to erase training-induced LTP. After a 6-day VR training protocol, each followed by CALI, a devaluation test was conducted to assess decision strategy. During the VR training, both CFL-SN and SN groups increased the number of lever presses (Fig. 4M), however, in the devaluation test the CFL-SN group showed a comparable number of lever presses in both the valued and devalued-conditions and a significantly smaller devaluation index than the SN group, indicating that their behavior became habitual (Fig. 4N

& O). Since, the two groups had a comparable level of lever pressing during training, this indicates that the acquisition of the lever pressing behavior itself does not require LTP in RSC-projecting ACC neurons (Fig. 4M). Additionally, CALI did not impact the preference and satiety for sucrose rewards in the devaluation processes, suggesting that the evaluation of reward values remains intact (Fig. S8).

Overall, the results suggest that training-induced excitatory synaptic potentiation in the RSC-projecting ACC neurons is required for the maintenance of the predominance of the goal-directed strategy, but if lost, the habitual strategy turns predominant. This is consistent with the observation that using VI training to establish habitual behavior, reduced the AMPA/NMDA ratio in this class of neuron (Fig. 4B–J).

RSC-projecting ACC neurons are active during reward acquisition of goal-directed behavior

Our data highlight the potentiation of synaptic transmission in a specific class of ACC neurons projecting to RSC, and their role in determining transitions in decision-making strategy. Given their importance and to better understand how these neurons influence behavior, we carried out Ca^{2+}-imaging in freely moving animals using a head-mounted fluorescent microscope[36].

RSC-projecting neurons were targeted by injecting a retrograde AAV vector encoding jGCaMP7s in the RSC (Fig. 5A). A subset of these neurons exhibited burst-like activity during rewarded nose pokes, whereas neurons showed minimal activity during lever pressing or non-rewarded nose pokes (Fig. 5B–D). The mean activity of neurons during rewarded nose pokes was significantly larger during VR training, and reduced following VR→VI switching (Fig. 5E). We classified RSC-projecting ACC neurons into reward-excited and -inhibited cells[37] (Fig. 5F–H; Methods). Significant changes in the fraction of each subclass were observed during the transition from goal-directed to habitual behavior (Fig. 5H). The proportion of reward-excited cells (Fig. 5H) as well as their activity (Fig. 5I) were significantly increased by the VR training and decreased after the VR→VI switching (Fig. 5I), consistent with the results of *ex vivo* electrophysiology. In contrast, the proportion of reward-inhibited

neurons (Fig. 5H) as well as their reward-evoked inhibition (Fig. 5J) decreased during the VR training but reverted following VR→VI switching. The increase in the reward-related excitation followed by goal-directed lever pressing and its reversion during habitual behavior align with the theory that the goal-directed strategy reflects current reward value in decision-making, whereas the habitual strategy makes decision without considering it. There was no significant correlation between changes in activity of these neurons and the execution index (Fig. S9). These results confirm that RSC-projecting ACC neurons are selectively potentiated during goal-directed behavior, which reverts upon a transition to habitual behavior.

Projection-specific plasticity of LOFC neurons after habit formation correlates with execution index

LOFC layer 5 neurons exhibited an increased AMPA/NMDA ratio during VR training, which was maintained in VR+VI_{High} but not in VR+VI_{Low} animals, thereby specifically correlating with the execution index after VR→VI switching (Fig. 2J, P). We therefore wanted to further explore how LOFC neurons may regulate changes in action execution accompanied by habit formation.

To achieve this, the retrograde labeling strategy was used to identify specific subclasses of LOFC neurons based on their projections. A retrograde AAV encoding tdTomato was injected separately into the CS, RSC, and BLA to label each subclass of projection neuron, which revealed LOFC neurons project to both CS and BLA, but not the RSC (Fig. 6A). The training protocol induced an increase in the AMPA/NMDA ratio, only in the CS, but not BLA-projecting neurons during the VR task (Fig. 6B & C). This ratio remained high in the VR+VI_{High} group, but not in the VR+VI_{Low} group (Fig. 6D & E). Among these two projection populations, only the CS-projecting neurons showed a significant correlation between AMPA/NMDA ratio and the execution index after VR+VI training, suggesting that the excitatory transmission of CS-projecting neurons determines the execution of habitual behavior (Fig. 6F & G).

Training-induced persistent synaptic potentiation of the CS-projecting LOFC neurons determines the execution of habitual behavior

To further investigate the causal relationship between the potentiation of CS-projecting LOFC neurons and action execution during habitual behavior, we expressed CFL-SN in the CS-projecting LOFC neurons and illuminated them after each VI session (days 10–13; Fig. 6H & I). After this period consisting of VI training with CALI, the average execution index and proportion of VR+VI_{High} mice in the CFL-SN group was significantly lower than the control (SN) group (Fig. 6J & K). Additionally, the effects of LTP erasure by CALI on the expression of habitual behavior was assessed by devaluation and omission tests. In the devaluation test, CALI in the CS-projecting LOFC neurons had no significant effect (Fig. 6L & M; Fig. S10A–C), likewise, in the omission test (Fig. S10A), CALI had no significant effect on the decay of lever press rate quantified by reduction in relative lever pressing rate (Fig. S10B, C). These results indicate that training-induced potentiation in CS-projecting LOFC neurons during habit formation, specifically determines the changes in execution level but not the decision strategy predominance.

Reward-related activity in the CS-projecting LOFC neurons after habit formation correlates with the execution index

To confirm this idea, activity of CS-projecting LOFC neurons during training by were recorded using Ca^{2+}-imaging (Fig. 7A). A significant response associated with both lever pressing and rewarded-nose pokes was observed (Fig. 7B & C). This contrasts with RSC projecting ACC neurons which responded only to the reward, but not the lever press (Fig. 5A–G). The overall activity of neurons during lever pressing was unchanged throughout the training session (Fig. 7D). In contrast, the activity during the rewarded-nose poke increased specifically in the VR+VI_{High} group, but not in the VR+VI_{Low} group (Fig. 7D). Responses to lever pressing and rewarded-nose pokes arose in two

non-overlapping populations of neurons (Fig. 7B & C), which were classified as lever press and reward neurons respectively (Fig. 7E; Methods). These neurons were mutually inhibited; the lever press neurons were inhibited during the firing of reward neurons and *vice versa* (Fig. 7F). Following a transition from goal-directed to habitual behavior, a significant reduction in the fraction of both classes of neurons were observed (Fig. 7E), with no correlation observed between the fractions of cells and execution index of each mouse (Fig. S11). Lever press cell activity was relatively stable across trials, with the exception of a small but statistically significant decrease during the VR training (Fig. 7G). There was no significant correlation between the lever press-related normalized AUC of the lever press cells and execution index of each mouse (Fig. 7H). In contrast, the average amplitude of the reward-related Ca^{2+} transients progressively increased across the training stages (Fig. 7I). Additionally, at the VI training stage, but not VR stage, the reward-related response had a positive correlation with the execution index of each mouse (Fig. 7J), suggestive of causal relationship between the reward-related activity of CS-projecting LOFC cells and execution index of habitual behavior. Taken together, reward-related activation of this subclass of neuron plays a pivotal role in sustaining a consistent level of execution during habit formation.

Discussion

Identifying the circuit mechanisms responsible for the transition from goal-directed to habitual behavior has proven difficult, primarily due to the lack of an appropriate behavior model where the habit formation can be induced at a precise time window. Our newly designed sequential task, combining VR and VI tasks allow us to accomplish this, since it comprises distinct goal-directed and habitual phases. Our task demonstrates that execution level is not always maintained following habit formation, but rather, within subject changes in execution level emerged during habit formation in parallel with the transition of decision-making strategy.

Our analysis revealed that the decision strategy, goal-directed versus habitual, and execution level are orthogonally represented in the brain. At the circuit level, the emergence of goal-directed behavior was mediated by potentiation of synaptic transmission in RSC-projecting ACC neurons. While, a transition away from goal-directed to habitual strategy was accompanied by a depotentiation of transmission in those same neurons. Finally, differences in execution level across subjects could be explained by the persistence/absence of synaptic potentiation in CS-projecting LOFC neurons. Together, these observations demonstrate that distinct populations of prefrontal cortical neurons govern predominance among decision strategies and execution of the selected habitual behavior.

The ACC is required for monitoring and prediction of an action-outcome relationship[38,39], by representing various reward-related information, such as reward value and cost of action[40–42]. Consistent with this, we found that chemogenetic manipulation of ACC activity bidirectionally regulated goal-directed/habitual decision strategy. Though recent studies have demonstrated that striatum-projecting ACC neurons are essential for cost-benefit assessment, but not goal-directed decision-making[43], while BLA-projecting neurons represent the innate value of reward and aversive signals[37,44], the mechanisms governing goal-directed decision-making have remained elusive. In this regard, this is the first demonstration that plasticity in RSC-projecting ACC neurons can dictate the goal-directed strategy for decision-making. Since both the ACC and RSC are crucial for learning and

memory of associative information, which is transferred from the hippocampus[35,45], it is conceivable that information of action-outcome association for goal-directed decision-making, which is transiently encoded in the hippocampus at earlier stages of training[46], may be transferred to the RSC-projecting ACC neurons.

Erasure of the VR training-induced LTP in the RSC-projecting ACC neurons did not impair learning of lever pressing behavior, while facilitating habit formation. As has been demonstrated by computational studies, two forms of reinforcement learning, model-based and model-free (or value-free) learning corresponding to goal-directed and habitual strategy respectively, in parallel progress during behavior acquision[2,47–49]. In a human study, habit formation is blocked while model-based learning is predominant, suggesting inhibitory mechanisms of the goal-directed strategy on the habitual strategy[50]. Accordingly, the results of the LTP erasure in RSC-projecting ACC neurons can be explained training-induced synaptic plasticity in the RSC-projecting ACC neurons serves to inhibit the habitual strategy.

In contrast to the ACC, although a number of studies implicate the OFC in reward-related decision-making, its functional significance in goal-directed decision-making is controversial; both activation and inhibition of the OFC disrupt flexibility of a behavior and shift it to habitual one[5,14]. In addition, consistent with our observations, previous studies have shown the OFC has no, or a limited contribution to behavior flexibility and defining decision strategies[51–53]. In line with this, comparative studies highlight the dissociative roles of the ACC and OFC in encoding outcome value; the OFC encodes outcome-related information during the Pavlovian (stimulus-outcome) conditioning rather than during the instrumental (action-outcome) conditioning, while the opposite contribution is found in the ACC[54, 55]. While possible causes for this inconsistency, such as heterogeneity of the OFC subdivisions and testing protocols, have been postulated[5,14,51–53], a specific role for the OFC in the goal-directed strategy is still to be fully elucidated.

Despite the functional differences between RSC-projecting ACC neurons and CS-projecting LOFC neurons, Ca^{2+}-imaging revealed that both neurons represent reward (outcome)-related information. Considering that *post hoc* LTP erasure was sufficient to affect the decision strategy and execution index, training-induced learning and updating of outcome-related association, rather than innate representation of reward value, may be critical for the effects of both neurons. As conceptualized by computational studies[2,47,48], rewarded experience updates the model for stimulus-response-outcome associations when using a model-based learning strategy. Hence, outcome-related information encoded in the RSC-projecting ACC neurons is likely to be utilized for updating the model of stimulus-response-outcome associations.

In contrast, outcome-related information encoded in the CS-projecting LOFC neurons is less likely to directly modulate the habitual strategy, since this strategy utilizes stimulus-response associations rather than action-outcome associations. Instead, it has been reported that the OFC is critical for contextual knowledge of outcome expectation, which hierarchically regulates performance of already acquired behaviors[56–58]. This function of the OFC has been mainly discussed in relation to flexible adaptation of behaviors according to context discrimination, where subjects learn distinct associations in a context-specific manner[5,58]; whereas, when subjects repeat stably rewarded experience in a specific context, this contextual stability is postulated to be able to facilitate execution of a habitual behavior, possibly through serving the contextual information as a promoter for the habitual (stimulus-response) decision-making[8,19,59]. Together, LTP in CS-projecting LOFC neurons could contribute to the association between rewarded experience and its context, thereby regulating the execution of lever pressing habit without affecting the stimulus-response associations.

Accumulating evidence from animal and clinical researches has shown that hyperactivity in the OFC-striatum pathway plays a pivotal role in the pathology of compulsivity[60–62]. Abnormally strong habit is considered as a possible cause of excessively repeated execution of certain behavior(s) shown in patients with compulsivity[63,64]. Symptom related cue-induced hyperactivity in the OFC and

increased functional connectivity between the OFC and striatum have been reported in patients with compulsive disorders[65–68]. While this present study focuses on habitual behavior in "healthy control" mice, the concept of LOFC-derived individuality in habit execution may be applicable to the pathological habitual behaviors such as obsessive-compulsive disorder, addiction and eating disorders.

In conclusion, the present study provides novel insights into neural mechanisms of behavior control underlying goal-directed to habitual transition. The spatiotemporally distinct online encoding of outcome-related information in the ACC and LOFC was required for blocking habit formation and maintenance of action execution, respectively (Fig. 7K). Our findings highlight the ACC-RSC pathway as a critical determinant of decision-making strategy, and further, shed light on the unexpected roles of LOFC-CS pathway on controlling habit execution.

Acknowledgments

We thank Drs. Dai Watanabe and Steven Middleton for their comments on the manuscript. This work was supported in part by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research JP22K15291 from the MEXT, Japan, Uehara Memorial Foundation, Kobayashi Foundation, Takeda Science Foundation and Astellas Foundation for Research on Metabolic Disorders (N.A.) and Grants-in-Aid for Scientific Research JP18H05434, JP20K21462 and JP22K21353 from the MEXT, Japan, The Uehara Memorial Foundation, The Naito Foundation, Research Foundation for Opto-Science and Technology, Novartis Foundation, and The Takeda Science Foundation, HFSP Research Grant RGP0022/2013, and JST CREST JPMJCR20E4 (Y.H.).

Declaration of interests

The authors declare no conflicts of interest.

Supplemental information

Document S1. Figure S1–12.

Table S1. Excel file containing statistical information.

Materials and Methods

Animals

All animal care and experimental procedures were conducted in accordance with the ethical guidelines of the Kyoto University Animal Research Committee. Male C57BL/6J mice (RRID: IMSR_JAX:000664) were purchased from Japan SLC (Shizuoka, Japan). Mice (6–24 weeks old) were housed at a constant ambient temperature of 24 ± 1°C on a 12-h light-dark cycle with access to food and water *ad libitum*. For the food restriction, mice were food-restricted (2–2.2 g/day) on weekdays (80–90% of the *ad libitum*-fed body weight). All behavioral experiments were randomized and videotaped.

Operant training

Operant training was carried out using an operant chamber (Med Associates, St Albans, VT) controlled by MED-PC software (version IV and V; Med Associates). Each session began with turning on the house light and lever presentation and ended when mice received the maximum number of reinforcers, or when 60 min had elapsed, unless otherwise described. For magazine training (2 or 3 (for Ca^{2+} recordings) sessions), food-restricted mice were placed in an operant chamber without lever presentation, after which 16 reinforcers (10 μL of 20% sucrose solution) were presented at 60-s intervals. For habituation to the lever, mice received three continuous reinforcement sessions, where each lever press was rewarded. The max number of reinforcers was increased across sessions (5, 15 and 30 reinforcers). Following CRF sessions, mice were trained with the VR or VI protocols. During VR training, an average of 10 (VR10) or 20 (VR20) presses were required for each reinforcer, whereas for VI training the reward was delivered on first press after an average interval time of either 30 (VI30) or 60 (VI60) sec following delivery of the previous reward. The sets of variables were fixed, but their order was changed across sessions. Mice that underwent the VR10 or VI30 schedule for the first two VR or VI sessions; thereafter were trained with the VR20 or VI60 schedule. When

mice were unable to obtain 10 reinforcers on the 4th VR20 or VI60 session, they were excluded from further experiments.

For the two-step paradigm, mice received three CRF, two VR10 and four VR20 sessions, which were followed by four VI60 sessions. The transition from VR20 to VI60 schedules were seamlessly carried out; without any cue or context changes. The changes in lever press rates between the last VR (fourth VR20) session and last VI60 session (fourth VI60 session) were indexed as follows;

Performance index = [(press rate on the VI) – (press rate on the VR)]/ [(press rate on the VI) + (press rate on the VR)].

For Fig. 7C, press rate in each VI60 session, was used instead of the last VI60 session

All lever presses, reward deliveries and nose pokes during the training were automatically recorded by the MED-PC software. The end of reward consumption was manually detected as the time at which mice start to move their head away from the reward port. For analyzing lever press bouts, presses with an inter-event interval of < 0.5 s were considered as one bout.

Devaluation test

After the last training session, devaluation tests were conducted over two consecutive days. Mice were allowed free access to food (the valued day) or a 20% sucrose solution (the devalued day) for 30 min. The order of the valued and devalued days was randomized. After the free access session, mice were placed in the operant chamber for 5 min where the lever was presented as usual, but never reinforced. The devaluation index was calculated as follows:

Devaluation index = [(lever presses on the valued day) – (lever presses on the devalued day)/ [(lever presses on the valued day) + (lever presses on the devalued day)].

Omission test

After the last training session, an omission test was conducted. During the 30-min omission session

the lever was presented as usual. When the lever was not pressed for 20 s, a reward (20% sucrose solution) was provided at the reward port, however when the lever was pressed a 20-s cooldown timer was reset.

For analysis, press rate was calculated every 5 min and divided by press rate in the last VI60 session. The time to reach 50% and 75% of the total press number was used to compare the decay of lever pressing activities.

Spatial discrimination learning and reversal learning

After completion of the two-step operant training, T-maze spatial discrimination training was performed as previously described[68]

. Food restriction was continued throughout all behavioral experiments. Mice were habituated with the T-maze apparatus (one start arm (30 × 10 cm), two goal arms (30 × 10 cm) and 30-cm-high surrounding walls) for 7 days and trained for the spatial discrimination for 14 days (5 trials/day). The rewarded arm (rewarded with 100 μL of sweetened milk) was randomly chosen and fixed during the 14-day training sessions. For reversal learning, the rewarded arm was reversed, and mice underwent 10 free-choice trials per day for 4 days.

Marble burying test

After completion of the T-maze experiments, the compulsive-like behavior was tested by the marble burying test. Each mouse was introduced in a novel plastic cage with 5-cm cage bedding (wood chips) and 15 glass marbles. The number of marbles which were covered with cage bedding more than 50% of their surface was counted across the 10-min session.

Stereotaxic surgery

Mice were anesthetized with isoflurane (1-2%) and fixed on a stereotaxic frame (Narishige, Tokyo,

Japan). Adeno-associated virus (AAV) vector (0.3 μL/side) was injected with stainless-steel cannula (30 G) at the following coordinates, unless otherwise described; ACC (AP 1.2 mm, ML ±0.5 mm, DV 1.7 mm), LOFC (AP 2.7 mm, ML ±1.7 mm, DV 2.5 mm), RSC (AP -1.0 mm, ML ±0.4 mm, DV 0.8 mm), CS (AP 1.2 mm, ML ±1.6 mm, DV 4.0 mm) and BLA (AP -1.5 mm, ML ±3.2 mm, DV 4.5 mm).

DREADD experiments

For DREADD experiments, the AAV vector expressing hM4Di (AAV-CaMKIIα-hM4Di-mCherry; Addgene #50477) or hM3Dq (AAV-CaMKIIα-hM3Dq-mCherry; Addgene #50476) was bilaterally injected into the ACC or lateral OFC. Mice were allowed to recover for at least one week prior to starting the operant training; so that at least 4-weeks had passed before CNO was injected.

For the devaluation tests, CNO (0.5 mg/kg; Cayman Chemical, MI, USA) or its vehicle (1% DMSO in saline) was injected prior to starting the free access session. For the cross-over design, mice were re-trained with one VR20 session after the first set of devaluation tests. The second set of devaluation tests were performed under the opposite drug condition. The order of the valued and devalued conditions was the same as in the first set. For the omission test, CNO or its vehicle were injected 20 min before starting the test.

In vivo Ca^{2+} imaging

Mice received unilateral injections of retroAAV-syn-jGCaMP7s-WPRE (Addgene #104487) in either the RSC or CS. To obtain a sufficient number of jGCaMP7s-positive cells, AAV was injected at two sites within the RSC or CS; RSC (AP -1.0 & -2.0 mm, ML ±0.4 mm, DV 0.8 mm) and CS (AP 1.2 & 0.2 mm, ML ±1.7 mm, DV 4.0 mm). Next a GRIN lens (1-mm diameter, 4-mm length; Inscopix, CA, USA) was placed in the ACC (AP 1.2 mm, ML ±0.8 mm, DV 0.5 mm) or LOFC (AP 2.7 mm, ML ±1.7 mm, DV 1.0 mm), after which mice were singly housed until experiments concluded.

At least 3 weeks post injection, a baseplate for the UCLA Miniscope[36] (version 4.4 purchased from OpenEphys) was fixed over the GRIN lens with dental cement. Lens implantation and baseplate mounting were performed with slight modifications to the method described at http://Miniscope.org. Operant training began at least 4 weeks after AAV injection. To enable mice with a Miniscope to enter the reward port, a wide custom-made nose poke hole was used. Mice underwent the two-step operant task with three magazine training sessions. For habituation to the Miniscope, mice were trained with a dummy scope after the second magazine training session. On the 3rd CRF, 6th VR and 4th VI sessions (the last session of each training schedule), calcium transients during the session were recorded at 30 Hz using the UCLA Miniscope DAQ software (version 1.10). Recordings of calcium transients and animal behaviors were started and finished simultaneously with the operant training sessions, by TTL signals from the MED-PC V software. Contrast and brightness of the videos were adjusted using FFmpeg and semi-automatically processed with the MiniAn pipeline[69] for removing sensor noise and background fluorescence, motion correction and source separation were performed by a constrained non-negative matrix factorization (CNMF) algorithm. The spatial footprint and calcium signal for each detected cell were manually verified to exclude false-positive detections.

Event-related Ca^{2+} activities

Following the MiniAn pipeline, output values of Ca^{2+} activity were normalized by their standard deviation (σ). For comparisons of the average Ca^{2+} responses (Fig. 5E, G, H & Fig. 7G, H, J), activity was Z-scored to highlight event-related activity changes from mean activity. Z-scores were calculated as follows; Z-score = $(C - \mu) / \sigma$, where C is an output value from the MiniAn pipeline and μ is mean of the overall output values. The time windows for lever pressing, non-rewarded nose pokes and rewarded nose poke events were determined as follows;

Lever press: A cluster of lever presses (inter-event interval (s) < 0.5 s) was treated as a single bout. The time window for analysis was set from 0.5 s before a bout, to 0.5 s after a bout. When mice did

a nose poke during this window, the event was excluded.

Non rewarded nose poke: When nose pokes were detected during unrewarded trials, these events were classified as non-rewarded nose pokes. The time window was set to ±0.5 s around the nose poke, if further nose pokes occurred within a 0.5 s interval, only the first one was included in analyses.

Rewarded nose poke: The first nose poke after reward delivery was considered as a rewarded nose poke. The time window from detection of rewarded nose pokes, to end of the reward consumption (manually detected as described above) was used for analysis. In the LOFC recording group, nine events (2 mice, CRF session) were excluded because the sensor detected objects other than mouse's nose (including the Miniscope, paw and tail). The pre-event period was set at 0.5 seconds prior to an event. Since the length of each event varied significantly, even within a single recording session, event-related Ca^{2+} activity was normalized by event length (time-normalized AUC). For excitation, the mean Ca^{2+} activity during events (σ) was divided by event length. For inhibition, the difference between the mean Ca^{2+} activities during events and pre-event periods, was divided by event length.

Analysis of Ca^{2+} transients in RSC-projecting ACC neurons

RSC-projecting ACC neurons were separated into three subclasses (reward-excited, reward-inhibited and non-responsive cells).

Reward-excited cells: Neurons which met the following three criteria were classified as reward-excited cells; (1) > 1 σ peak amplitude in at least 20% of rewarded nose poke events, (2) mean amplitude during rewarded nose pokes was larger than that during pre-nose poke periods, (3) showing multiple peaks during rewarded nose pokes (Fig. 5D). Since, typical rewarded nose pokes lasted 2-20 s, the criteria (2) and (3) were used to exclude a subset of neurons which showed non-event-related episodic Ca^{2+} activity.

Reward-inhibited cells: Neurons that were not classified as reared-excited cells were further tested for their decreases in Ca^{2+} activities. The minimum value of Ca^{2+} activity during each rewarded nose

poke was compared with the mean amplitude during the corresponding pre-nose poke period using a paired *t* test. When there was a significant decrease in Ca^{2+} activity, the neuron was classified as a reward-inhibited cell.

For further statistical analyses, excitatory/inhibitory magnitude of each reward-excited/inhibited cell was scored as a population-weighted response (Fig. 5I & J), which was calculated by normalized AUC of rewarded nose pokes, multiplied by the fraction of reward-excited/inhibited cells in each mouse[70].

Classification of activity of CS-projecting LOFC neurons

CS-projecting LOFC neurons were separated into three subclasses (reward, lever press and not responsible cells), as follows.

First, normalized AUCs of Ca^{2+} activities for reward and lever press events were calculated. Among neurons recorded in the VR session, time-normalized AUCs for each event were Z-scored, so that features of response selectivity could be compared. To this end, neurons recorded in the VR session were clustered into 5 clusters by the K-means+ method (Python with scikit-learn package). The number of clusters was determined by the elbow method. Then, to apply this classification to recordings in the CRF and VI sessions, we made a logistic regression model for classification of neurons using time-normalized AUCs datasets. For training, 75% of the datasets in the VR session were used. The accuracy of the model tested with the remaining 25% of the datasets was 0.89.

Using the classification, neurons were separated into 5 groups: with high and moderate preference for lever pressing, high and moderate preference for reward and non-responsive for both events (Fig. S12). Since the numbers of high preference neurons were small, high and moderate preference groups for each event were combined. For each event-related neurons normalized AUC and population weighted normalized AUC were calculated as described above.

CALI experiments

pAAV-CaMKIIα-CFL-SN and pAAV-CaMKIIα-SN were constructed by replacement of the hM4Di-mCherry gene in the pAAV-CaMKIIα-hM4Di-mCherry (Addgene #50477) virus, to the CFL-SN and SN genes, respectively. Retrograde AAV vectors for CFL-SN and SN expression were produced as previously described[68].

Mice received bilateral injections of retroAAV-CaMKIIα-CFL-SN or retroAAV-CaMKIIα -SN in the RSC or CS. Fiber optic cannulae, constructed from ceramic ferrules (1.25 mm diameter; CFLC270-10; Thorlabs NJ, USA) and plastic optic fibers (250 μm diameter; CK-10; Mitsubishi Rayon, Tokyo, Japan), were implanted in the ACC (AP 1.2 mm, ML ±0.5 mm, DV 1.0 mm, angled 10° relative to the ML axis) or LOFC (AP 2.7 mm, ML ±1.7 mm, DV 2.5 mm). 595 nm laser light intensity from the tip of the fiber optic cannula was adjusted to 1–2 mW. Operant training commenced at least 1 week post-surgery, so that at least 3 weeks had elapsed since injection of virus.

For CALI in the RSC-projecting ACC neurons, mice received bilateral illumination of the ACC for 1 min, within 5 min after VR training had ended. For CALI in the CS-projecting LOFC neurons, mice received bilateral illumination of the LOFC for 1 min, within 5 min after VI training had ended.

Histology

After completing all behavioral experiments, mice were deeply anesthetized with isoflurane and transcardially perfused with PBS followed by 4% paraformaldehyde. Whole brain samples were further fixed in 4% paraformaldehyde for 30 min at 4°C and incubated in 15% sucrose/PBS solution overnight at 4°C. Coronal brain sections (30-μm thick) were prepared with a cryostat (Leica CM1860; Leica Biosystems, Nußloch, Germany). Signals from fluorescent proteins were visualized by a confocal microscopy system (FLUOVIEW FV1200, Olympus, Tokyo, Japan). For visualizing jGCaMP7s, sections were stained with rabbit anti-GFP antibody (1:500; A11122, lot# 2659306; Thermo Fisher Scientific, MA, USA) followed by AlexaFluor 594 conjugated secondary antibodies

(1:200; Cell Signaling Technology, MA, USA).

Electrophysiology recordings

Electrophysiology recordings were performed as previously described[68,71] with minor modifications. For the control group, mice were food restricted at least 2 weeks, but did not receive operant training. In the projection-specific subpopulation groups, mice were injected with retroAAV-CAG-tdTomato (Addgene #59462) at least 3 weeks before recording.

Within 30-min after the last operant training session, mice were deeply anesthetized with isoflurane and decapitated. Coronal brain slices (200-μm thick) were prepared with a vibratome (VT1000S, Leica, Wetzlar, Germany) filled with ice-cold cutting solution (in mM: 120 NMDG-Cl, 2.5 KCl, 26 $NaHCO_3$, 1.25 NaH_2PO_4, 0.5 $CaCl_2$, 7 $MgCl_2$, 15 D-glucose, and 1.3 ascorbic acid, pH 7.2). Slices were recovered in ACSF (composition in mM: 124 NaCl, 3 KCl, 26 $NaHCO_3$, 1 NaH_2PO_4, 2.4 $CaCl_2$, 1.2 $MgCl_2$, and 10 D-glucose, pH 7.3, oxygenated with 95%O_2/5% CO_2) at 32°C for at least 1 h before recording. After recovery, individual slices were transferred to a recording chamber with continuous perfusion of oxygenated ACSF, heated to 27 ± 1°C.

Signals were passed to an EPC9 amplifier (HEKA, Pfalz, Germany) and data were recorded using Patchmaster software (HEKA). Individual neurons were visualized with a microscope equipped with a 40 × water-immersion objective lens (Carl Zeiss, Jena, Germany) and a CCD camera. The resistance of electrodes was 3-7 MΩ when filled with the internal solution (in mM: 120 $CsMeSO_4$, 15 CsCl, 8 NaCl, 10 HEPES, 2 Mg-ATP, 0.3 Na-GTP, 0.2 EGTA, 10 TEA-Cl, and 5 QX-314, pH 7.3 adjusted with CsOH). The series resistance was compensated by 70% and maintained at 35 MΩ. Additionally, a glass stimulating electrode filled with ACSF was placed near the recording cells.

To record EPSCs, membrane potentials were held at -70 mV and a $GABA_A$ antagonist (20 μM bicuculline) was continuously perfused. AMPA-mediated evoked EPSCs (eEPSCs) and mixed AMPA and NMDA-mediated eEPSCs were evoked by stimulation at -70 mV and +40 mV, respectively.

NMDA-mediated eEPSC amplitude was determined as the average amplitude between 45 and 55 ms post-stimulation. The average of 5 consecutive AMPA- and NMDA-mediated eEPSC measurements was used for analyzing AMPA/NMDA ratios. AMPA receptor (AMPAR)-mediated eEPSCs were normalized by NMDA receptor (NMDAR)-mediated eEPSCs. During paired pulse ratio experiments, electrical stimuli were delivered with a 30-ms interval. The average of 5 consecutive PPRs was used for analysis.

When recording asynchronous EPSCs, $CaCl_2$ in the ACSF was substituted with 6 mM $SrCl_2$. Asynchronous events evoked by electrical stimulation were recorded 10 times from each neuron and analyzed by Minianalysis software (SynaptoSoft, Decatur, GA) and Win EDR 3.7.3 (University of Strathclyde). Events with amplitudes larger than 15 pA were used for analysis.

Statistical analyses

Data are presented as the mean ± standard error of mean (S.E.M) unless otherwise stated. Box and whisker plots show the maximum, 75th percentile, median, 25th percentile and minimum values. Statistical analyses were performed with GraphPad Prism 8 (GraphPad, CA, USA), P values < 0.05 considered significant. All statistical methods, sample sizes and detailed statistics are shown in Supplementary Table 1.

REFERENCES

1. Rangel, A., Camerer, C., & Montague, P. A. (2008). A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci, 9, 545–556. https://doi.org/10.1038/nrn2357
2. Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. Neuron, 80(2), 312-25. https://doi.org/10.1016/j.neuron.2013.09.007
3. Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. Philosophical Transactions of the Royal Society, B308, 67–78. http://doi.org/10.1098/rstb.1985.0010
4. Du, Y., Krakauer, J. W., & Haith, A. M. (2022). The relationship between habits and motor skills in humans. Trends Cogn Sci, 26(5), 371-387. https://doi.org/10.1016/j.tics.2022.02.004
5. Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nat Commun, 4, 2264. https://doi.org/10.1038/ncomms3264
6. Gardner, B. (2015). A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. Health Psychol Rev. 9(3), 277-95. doi: 10.1080/17437199.2013.876238
7. Wood, W., & Rünger, D. (2016). Psychology of Habit. Annu Rev Psychol, 67, 289-314. https://doi.org/10.1146/annurev-psych-122414-033417
8. Buabang EK, Donegan KR, Rafei P, Gillan CM. (2025). Leveraging cognitive neuroscience for making and breaking real-world habits. Trends Cogn Sci. 29(1), 41-59. doi: 10.1016/j.tics.2024.10.006
9. Courtin J, Bitterman Y, Müller S, Hinz J, Hagihara KM, Müller C, Lüthi A. (2022) A neuronal mechanism for motivational control of behavior. Science. 375(6576), eabg7277. doi: 10.1126/science.abg7277
10. Gardner, B., Phillips, L. A., & Judah, G. (2016). Habitual instigation and habitual execution: Definition, measurement, and effects on behaviour frequency. Br J Health Psychol, 21(3), 613-30. https://doi.org/10.1111/bjhp.12189
11. Balleine, B. W., & Dezfouli, A. (2019). Hierarchical Action Control: Adaptive Collaboration Between Actions and Habits. Front Psychol, 10, 2735. https://doi.org/10.3389/fpsyg.2019.02735
12. Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology, 35(1), 48-69. https://doi.org/10.1038/npp.2009.116
13. O'Hare, J. K., Ade, K. K., Sukharnikova, T., Van Hooser, S. D., Palmeri, M. L., Yin, H. H., & Calakos, N. (2016). Pathway-Specific Striatal Substrates for Habitual Behavior. Neuron, 89(3), 472-9. https://doi.org/10.1016/j.neuron.2015.12.036
14. Duan, L. Y., Horst, N. K., Cranmore, S. A. W., Horiguchi, N., Cardinal, R. N., Roberts, A. C., & Robbins, T. W. (2021). Controlling one's world: Identification of sub-regions of primate PFC underlying goal-directed behavior. Neuron, 109(15), 2485-2498.e5. https://doi.org/10.1016/j.neuron.2021.05.010
15. Yamada, K., & Toda, K. (2023). Habit formation viewed as structural change in the behavioral network. Commun Biol, 6, 303. https://doi.org/10.1038/s42003-023-02968-9
16. Hilário, M. R., Clouse, E., Yin, H. H., & Costa, R. M. (2007). Endocannabinoid signaling is

critical for habit formation. Front Integr Neurosci, 1, 6. https://doi.org/10.3389/neuro.07.006.2007

17. Derusso, A. L., Fan, D., Gupta, J., Shelest, O., Costa, R. M., & Yin, H. H. (2010). Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. Front Integr Neurosci, 4, 17. https://doi.org/10.3389/fnint.2010.00017

18. Baker, M., Kang, S., Hong, S. I., Song, M., Yang, M. A., Peyton, L., Essa, H., Lee, S. W., & Choi, D. S. (2023). External globus pallidus input to the dorsal striatum regulates habitual seeking behavior in male mice. Nat Commun, 14(1), 4085. https://doi.org/10.1038/s41467-023-24342-5

19. Bouton, M.E. (2021). Context, attention, and the switch between habit and goal-direction in behavior. Learn Behav 49, 349–362. https://doi.org/10.3758/s13420-021-00488-z

20. Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav Brain Res, 166(2), 189-96. https://doi.org/10.1016/j.bbr.2005.07.012

21. Seiler, J. L., Cosme, C. V., Sherathiya, V. N., Schaid, M. D., Bianco, J. M., Bridgemohan, A. S., & Lerner, T. N. (2022). Dopamine signaling in the dorsomedial striatum promotes compulsive behavior. Curr Biol, 32(5), 1175-1188.e5. https://doi.org/10.1016/j.cub.2021.12.046

22. Gremel, C. M., Chancey, J. H., Atwood, B. K., Luo, G., Neve, R., Ramakrishnan, C., Deisseroth, K., Lovinger, D. M., & Costa, R. M. (2016). Endocannabinoid Modulation of Orbitostriatal Circuits Gates Habit Formation. Neuron, 90(6), 1312-1324. https://doi.org/10.1016/j.neuron.2016.04.043

23. Rossi, M. A., & Yin, H. H. (2012). Methods for studying habitual behavior in mice. Curr Protoc Neurosci, Chapter 8, Unit 8.29. https://doi.org/10.1002/0471142301.ns0829s59

24. Rushworth, M., & Behrens, T. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci, 11, 389–397. https://doi.org/10.1038/nn2066

25. Takahashi, T., Svoboda, K., & Malinow, R. (2003). Experience strengthening transmission by driving AMPA receptors into synapses. Science, 299(5612), 1585-8. https://doi.org/10.1126/science.1079607

26. Watt AJ, Sjöström PJ, Häusser M, Nelson SB, Turrigiano GG. (2004). A proportional but slower NMDA potentiation follows AMPA potentiation in LTP. Nat Neurosci. 7(5), 518-24. doi: 10.1038/nn1220

27. Zhang, S., Xu, M., Chang, W. C., Ma, C., Hoang Do, J. P., Jeong, D., Lei, T., Fan, J. L., & Dan, Y. (2016). Organization of long-range inputs and outputs of frontal cortex for top-down control. Nat Neurosci, 19(12), 1733-1742. https://doi.org/10.1038/nn.4417

28. Yao, Z., van Velthoven, C. T. J., Nguyen, T. N., Goldy, J., Sedeno-Cortes, A. E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., Ding, S. L., Fong, O., Garren, E., Glandon, A., Gouwens, N. W., Gray, J., Graybuck, L. T., Hawrylycz, M. J., Hirschstein, D., Kroll, M., Lathia, K., Lee, C., Levi, B., McMillen, D., Mok, S., Pham, T., Ren, Q., Rimorin, C., Shapovalova, N., Sulc, J., Sunkin, S. M., Tieu, M., Torkelson, A., Tung, H., Ward, K., Dee, N., Smith, K. A., Tasic, B., & Zeng, H. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell, 184(12), 3222-3241.e26. https://doi.org/10.1016/j.cell.2021.05.030

29. Gabbott, P. L., Warner, T. A., Jays, P. R., Salway, P., & Busby, S. J. (2005). Prefrontal cortex in the rat: projections to subcortical autonomic, motor, and limbic centers. J Comp Neurol, 492(2), 145-77. https://doi.org/10.1002/cne.20738

30. Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. Nat Rev Neurosci, 7(6), 464-76. https://doi.org/10.1038/nrn1919

31. Todd, T. P., Fournier, D. I., & Bucci, D. J. (2019). Retrosplenial cortex and its role in cue-specific learning and memory. Neurosci Biobehav Rev, 107, 713-728. https://doi.org/10.1016/j.neubiorev.2019.04.016

32. Shi, W., Xue, M., Wu, F., Fan, K., Chen, Q. Y., Xu, F., Li, X. H., Bi, G. Q., Lu, J. S., & Zhuo, M. (2022). Whole-brain mapping of efferent projections of the anterior cingulate cortex in adult male mice. Mol Pain, 18, 17448069221094529. https://doi.org/10.1177/17448069221094529

33. Wassum, K. M. (2022). Amygdala-cortical collaboration in reward learning and decision making. eLife, 11, e80926. https://doi.org/10.7554/eLife.80926

34. Bosch, M., Castro, J., Saneyoshi, T., Matsuno, H., Sur, M., & Hayashi, Y. (2014). Structural and molecular remodeling of dendritic spine substructures during long-term potentiation. Neuron, 82(2), 444-59. https://doi.org/10.1016/j.neuron.2014.03.021

35. Goto, A., Bota, A., Miya, K., Wang, J., Tsukamoto, S., Jiang, X., Hirai, D., Murayama, M., Matsuda, T., McHugh, T. J., Nagai, T., & Hayashi, Y. (2021). Stepwise synaptic plasticity events drive the early phase of memory consolidation. Science (New York, N.Y.), 374(6569), 857–863. https://doi.org/10.1126/science.abj9195

36. Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B., Veshkini, M., La-Vu, M., Lou, J., Flores, S. E., Kim, I., Sano, Y., Zhou, M., Baumgaertel, K., Lavi, A., Kamata, M., Tuszynski, M., Mayford, M., Golshani, P., & Silva, A. J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. Nature, 534(7605), 115-8. https://doi.org/10.1038/nature17955

37. Yuan, Z., Qi, Z., Wang, R., Cui, Y., An, S., Wu, G., Feng, Q., Lin, R., Dai, R., Li, A., Gong, H., Luo, Q., Fu, L., & Luo, M. (2023). A corticoamygdalar pathway controls reward devaluation and depression using dynamic inhibition code. Neuron, 111(23), 3837-3853.e5. https://doi.org/10.1016/j.neuron.2023.09.008

38. Kennerley, S., Walton, M., Behrens, T. E., Buckley, M. J., & Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. Nat Neurosci, 9, 940–947. https://doi.org/10.1038/nn1724

39. Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R. F., Dayan, P., & Costa, R. M. (2021). The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection. Neuron, 109(1), 149-163.e7. https://doi.org/10.1016/j.neuron.2020.10.041

40. Rudebeck, P., Walton, M., Smyth, A., Bannerman, D., & Rushworth, M. (2006). Separate neural pathways process different decision costs. Nat Neurosci, 9, 1161–1168. https://doi.org/10.1038/nn1756

41. Monosov, I. E. (2017). Anterior cingulate is a source of valence-specific information about value and uncertainty. Nat Commun, 8, 134. https://doi.org/10.1038/s41467-017-00126-6

42. Fetcho, R. N., Parekh, P. K., Chou, J., Kenwood, M., Chalençon, L., Estrin, D. J., Johnson, M., & Liston, C. (2024). A stress-sensitive frontostriatal circuit supporting effortful reward-seeking behavior. Neuron, 112(3), 473-487.e4. https://doi.org/10.1016/j.neuron.2024.01.012

43. Ulloa Severino, F. P., Lawal, O. O., Sakers, K., Saghy, F., Galván-Arevalo, J., Lee, A. T., Noguchi, J., Gornet, M. K., Peixoto, R. T., & Kramer, F. (2023). Training-induced circuit-specific excitatory synaptogenesis in mice is required for effort control. Nat Commun, 14, 5522. https://doi.org/10.1038/s41467-023-38994-5

44. Jhang, J., Lee, H., Kang, M. S., Lee, H. S., Cho, H., Ryu, C., Kim, D., & Lee, S. H. (2018). Anterior cingulate cortex and its input to the basolateral amygdala control innate fear response. Nat Commun, 9, 2744. https://doi.org/10.1038/s41467-018-05030-3

45. de Sousa, A. F., Cowansage, K. K., Zutshi, I., Cardozo, L. M., Yoo, E. J., Leutgeb, S., & Mayford, M. (2019). Optogenetic reactivation of memory ensembles in the retrosplenial cortex induces systems consolidation. Proc Natl Acad Sci U S A, 116(17), 8576-8581. https://doi.org/10.1073/pnas.1819895116

46. Bradfield, L. A., Leung, B. K., Boldt, S., & Balleine, B. W. (2020). Goal-directed actions transiently depend on dorsal hippocampus. Nat Neurosci, 23, 1194–1197. https://doi.org/10.1038/s41593-020-0635-6

47. Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron, 66(4), 585–595. https://doi.org/10.1016/j.neuron.2010.04.016

48. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron, 69(6), 1204-15. https://doi.org/10.1016/j.neuron.2011.02.027

49. Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values. Psychological review, 126(2), 292–311. https://doi.org/10.1037/rev0000120

50. Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. Cogn Affect Behav Neurosci, 15(3), 523-36. https://doi.org/10.3758/s13415-015-0347-6

51. Ostlund, S. B., & Balleine, B. W. (2007). Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. J Neurosci, 27(18), 4819-25. https://doi.org/10.1523/JNEUROSCI.5443-06.2007

52. Parkes, S. L., Ravassard, P. M., Cerpa, J. C., Wolff, M., Ferreira, G., & Coutureau, E. (2018). Insular and Ventrolateral Orbitofrontal Cortices Differentially Contribute to Goal-Directed Behavior in Rodents. Cereb Cortex, 28(7), 2313-2325. https://doi.org/10.1093/cercor/bhx121

53. Panayi, M. C., & Killcross, S. (2018). Functional heterogeneity within the rodent lateral orbitofrontal cortex dissociates outcome devaluation and reversal learning deficits. eLife, 7, e37357. https://doi.org/10.7554/eLife.37357

54. Luk, C. H., & Wallis, J. D. (2013). Choice coding in frontal cortex during stimulus-guided or action-guided decision-making. J Neurosci, 33(5), 1864-71. https://doi.org/10.1523/JNEUROSCI.4929-12.2013

55. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. Neuron, 81(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

56. Farovik, A., Place, R. J., McKenzie, S., Porter, B., Munro, C. E., & Eichenbaum, H. (2015). Orbitofrontal cortex encodes memories within value-based schemas and represents contexts that guide memory retrieval. J Neurosci, 35(21), 8333–8344. https://doi.org/10.1523/JNEUROSCI.0134-15.2015

57. Peterson, S., Chavira, J., Garcia Arango, J. A., Seamans, D., Cimino, E. D., & Keiflin, R. (2024). Partially dissociable roles of the orbitofrontal cortex and dorsal hippocampus in context-dependent hierarchical associations. Current biology, 34(23), 5532–5545.e3. https://doi.org/10.1016/j.cub.2024.10.049

58. Thrailkill, E. A., & Bouton, M. E. (2015). Contextual control of instrumental actions and habits. Journal of experimental psychology. Animal learning and cognition, 41(1), 69–80. https://doi.org/10.1037/xan0000045

59. Ahmari, S. E., Spellman, T., Douglass, N. L., Kheirbek, M. A., Simpson, H. B., Deisseroth, K., Gordon, J. A., & Hen, R. (2013). Repeated cortico-striatal stimulation generates persistent OCD-like behavior. *Science* (New York, N.Y.), **340**, 1234–1239. https://doi.org/10.1126/science.1234733

60. Abe, Y., Sakai, Y., Nishida, S., Nakamae, T., Yamada, K., Fukui, K., & Narumoto, J. (2015). Hyper-influence of the orbitofrontal cortex over the ventral striatum in obsessive-compulsive disorder. *European neuropsychopharmacology*, **25**, 1898–1905. https://doi.org/10.1016/j.euroneuro.2015.08.017

61. Pascoli, V., Hiver, A., Van Zessen, R., Loureiro, M., Achargui, R., Harada, M., Flakowski, J., & Lüscher, C. (2018). Stochastic synaptic plasticity underlying compulsion in a model of addiction. *Nature*, **564**, 366–371. https://doi.org/10.1038/s41586-018-0789-4

62. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. eLife, 5, e11305. https://doi.org/10.7554/eLife.11305

63. Robbins, T. W., Vaghi, M. M., & Banca, P. (2019). Obsessive-Compulsive Disorder: Puzzles and Prospects. Neuron, 102(1), 27-47. https://doi.org/10.1016/j.neuron.2019.01.045

64. Filbey, F. M., Claus, E., Audette, A. R., Niculescu, M., Banich, M. T., Tanabe, J., Du, Y. P., & Hutchison, K. E. (2008). Exposure to the taste of alcohol elicits activation of the mesocorticolimbic neurocircuitry. Neuropsychopharmacology, 33(6), 1391-401. https://doi.org/10.1038/sj.npp.1301513

65. Rotge, J. Y., Guehl, D., Dilharreguy, B., Cuny, E., Tignol, J., Bioulac, B., Allard, M., Burbaud, P., & Aouizerate, B. (2008). Provocation of obsessive-compulsive symptoms: a quantitative voxel-based meta-analysis of functional neuroimaging studies. J Psychiatry Neurosci, 33(5), 405-12

66. Schienle, A., Schäfer, A., Hermann, A., & Vaitl, D. (2009). Binge-eating disorder: reward sensitivity and brain activation to images of food. Biol Psychiatry, 65(8), 654-61. https://doi.org/10.1016/j.biopsych.2008.09.028

67. Wang, A. R., Kuijper, F. M., Barbosa, D. A. N., Hagan, K. E., Lee, E., Tong, E., Choi, E. Y., McNab, J. A., Bohon, C., & Halpern, C. H. (2023). Human habit neural circuitry may be perturbed in eating disorders. Sci Transl Med, 15(689), eabo4919. https://doi.org/10.1126/scitranslmed.abo4919

68. Asaoka, N., Ibi, M., Hatakama, H., Nagaoka, K., Iwata, K., Matsumoto, M., Katsuyama, M., Kaneko, S., & Yabe-Nishimura, C. (2021). NOX1/NADPH Oxidase Promotes Synaptic Facilitation Induced by Repeated D2 Receptor Stimulation: Involvement in Behavioral Repetition. J Neurosci, 41(12), 2780-2794. https://doi.org/10.1523/JNEUROSCI.0586-20.2021

69. Dong, Z., Mau, W., Feng, Y., Pennington, Z. T., Chen, L., Zaki, Y., Rajan, K., Shuman, T., Aharoni, D., & Cai, D. J. (2022). Minian, an open-source miniscope analysis pipeline. eLife, 11, e70661. https://doi.org/10.7554/eLife.70661

70. Grove, J. C. R., Gray, L. A., La Santa Medina, N., Sivakumar, N., Ahn, J. S., Corpuz, T. V., Berke, J. D., Kreitzer, A. C., & Knight, Z. A. (2022). Dopamine subsystems that track internal states. Nature, 608(7922), 374-380. https://doi.org/10.1038/s41586-022-05054-2

71. Hatakama, H., Asaoka, N., Nagayasu, K., Shirakawa, H., & Kaneko, S. (2024). Amelioration of

obsessive-compulsive disorder by intracellular acidification of cortical neurons with a proton pump inhibitor. Transl Psychiatry, 14(1), 27. https://doi.org/10.1038/s41398-024-00354-5

FIGURE LEGENDS

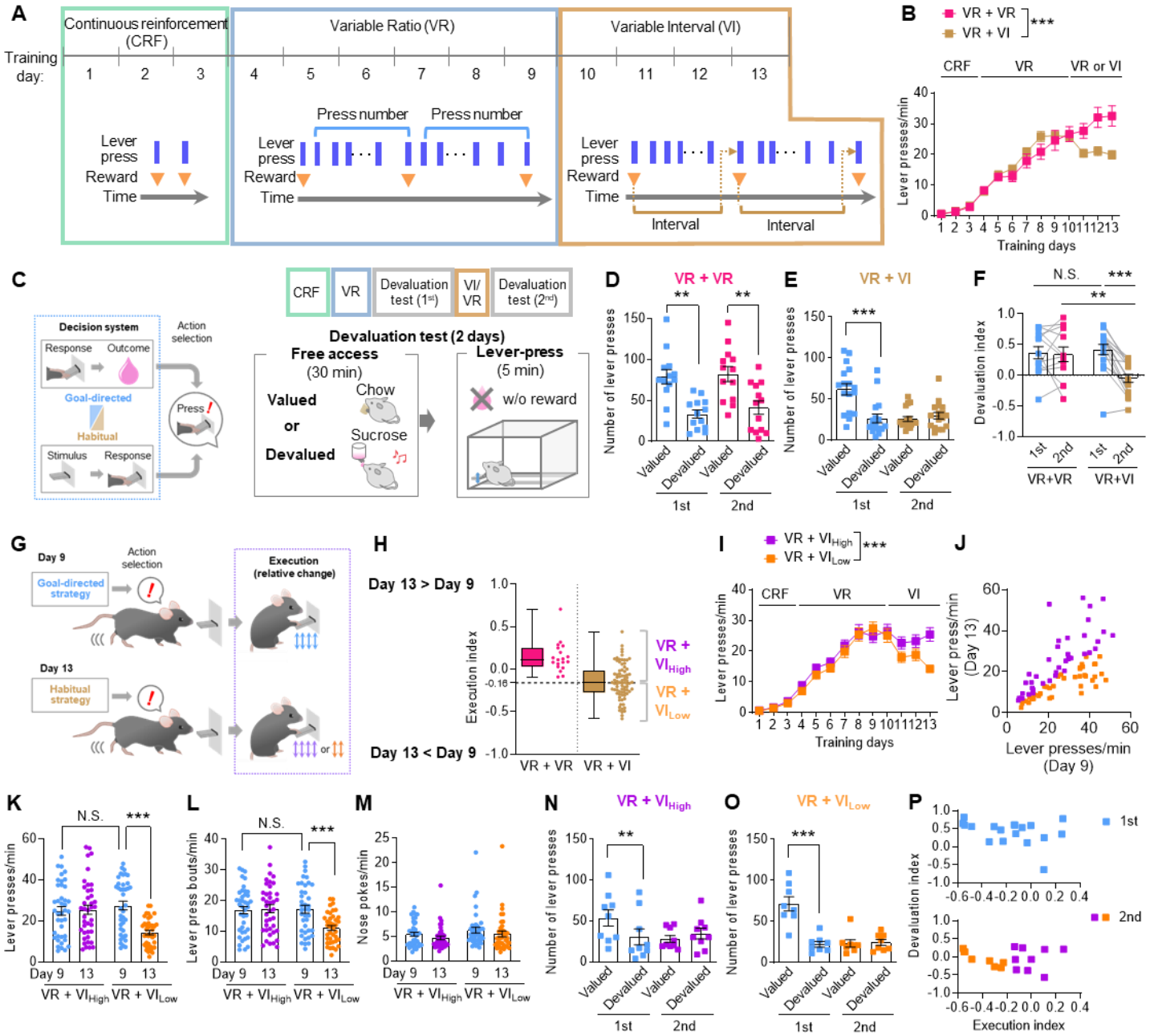


Figure 1

Individual differences in the execution of lever pressing during the two-step training

(A) Schematic diagram of the two-step operant training protocol. On day 10, the training schedule was seamlessly switched from a variable ratio (VR) to a variable interval (VI) task (VR+VI). Mice without the VR→VI switching (VR+VR) are considered as the control group. **(B)** Transitions of lever pressing rate in the two-step training. **(C)** Schematic diagram of the devaluation test for evaluation of decision strategy. Mice underwent the first devaluation test after the 6th VR session. Then, after the

4th day of VI or VR training, the second devaluation test was performed. On each test day, mice were fed with regular chow, or sucrose solution which mice learned to associate with lever pressing. With free-access to sucrose solution, its relative value declined (devalued condition). After 30-min free-access, mice were placed in the operant chamber and allowed to press the lever for 5 min, during which no sucrose was delivered. **(D, E)** Numbers of lever presses on each test day in the VR+VR **(D)** and VR+VI **(E)** groups. **(F)** Within-subject differences in lever pressing between valued and devalued conditions, are shown as the devaluation index. **(G)** To assess the impact of habit formation on behavioral execution, lever pressing activities before and after habit formation were compared in a within-subject manner. **(H)** Within-subject differences in the lever press rate between days 9 & 13 are plotted as the execution index. The VR+VI group was subdivided into the $VR+VI_{High}$ and $VR+VI_{Low}$ groups by the median value of the execution index. **(I)** Lever press rate in the $VR+VI_{High}$ and $VR+VI_{Low}$ groups. **(J)** Single session lever press rate on days 9 & 13 in the $VR+VI_{High}$ and $VR+VI_{Low}$ groups. **(K)** Comparison of the lever press rate on days 9 & 13, reveal a significant difference in the distribution only on day 13. **(L)** The number of lever press bouts on days 9 & 13. VR→VI switching reduced both initiation of lever pressing and bout size in the $VR+VI_{Low}$ group. **(M)** Frequency of nose pokes was not affected by VR→VI switching. **(M , O)** Number of lever presses in each devaluation test in the VR+VI group **(E)** are separated into the $VR+VI_{High}$ and $VR+VI_{Low}$ groups. **(P)** Within-subject differences in lever pressing between the valued and devalued conditions are shown as the devaluation index. Relations between the execution index and the devaluation index in the first and second devaluation tests are plotted in XY-planes. $^{**}P < 0.01$ and $^{***}P < 0.001$; N.S., not significant.

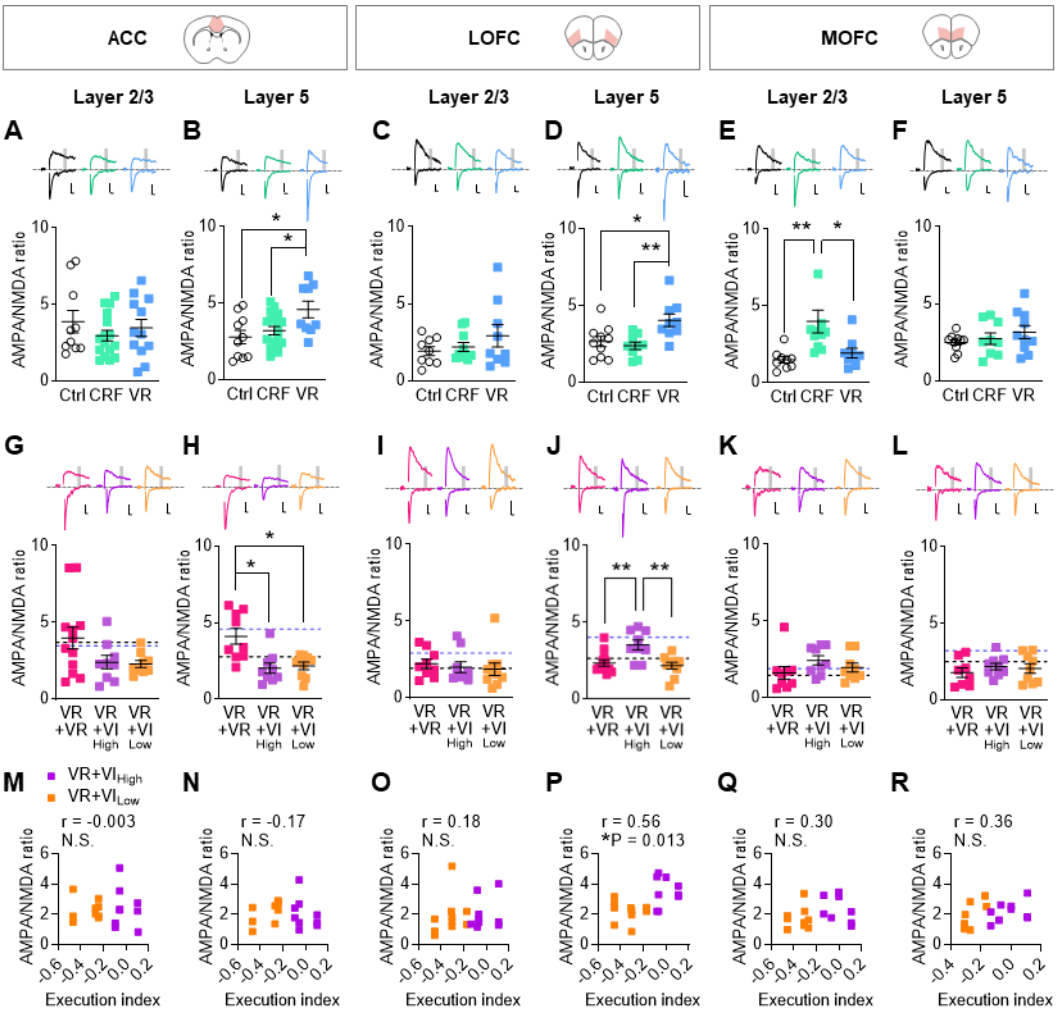


Figure 2

Cortical region-specific plasticity is induced at each training stage

(A–F) Acute brain slices of the anterior cingulate cortex (ACC), lateral orbitofrontal cortex (OFC) and medial OFC were prepared within 30 min after the 3rd CRF and 6th VR training. Control (Ctrl) mice received food restriction but were not trained. AMPA/NMDA ratios were recorded from pyramidal neurons in the ACC layer 2/3 **(A)**, ACC layer 5 **(B)**, LOFC layer 2/3 **(C)**, LOFC layer 5 **(D)**, medial OFC layer 2/3 **(E)** and medial OFC layer 5 **(F)**. **(G–L)** Acute brain slices of the ACC, LOFC and medial OFC were prepared within 30 min after the 4th VI training session. The VR+VR group was introduced as a control matched for total training days. AMPA/NMDA ratios were recorded from pyramidal neurons in the ACC layer 2/3 **(G)**, ACC layer 5 **(H)**, LOFC layer 2/3 **(I)**, LOFC layer 5 **(J)**, medial OFC layer 2/3 **(K)** and medial OFC layer 5 **(L)**. The black and blue dashed lines show the mean value of the Ctrl and VR groups in **(A–F)** respectively. **(M–R)** Correlation between the

execution index and AMPA/NMDA ratios in the VR+VI_{High} and VR+VI_{Low} groups. *$P < 0.05$; **$P < 0.01$.

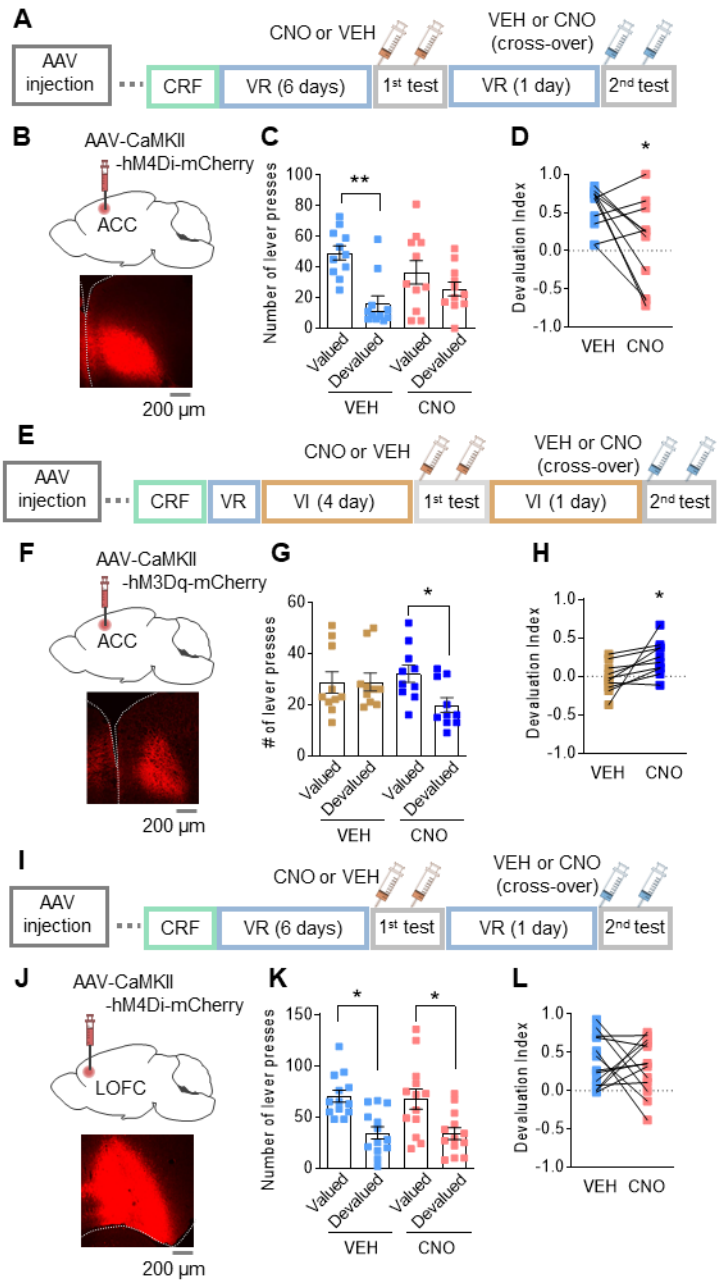


Figure 3

Chemogenetic manipulation of the ACC but not LOFC neurons affects habitual decision-making

(A) Mice received injection with AAV-CaMKII-hM4Di-mCherry in the ACC. On the next two days after the 6th VR training, decision-making strategy was assessed by the devaluation test (first test). Before the free-access sessions, clozapine-N-oxide (CNO; 0.5 mg/kg) or its vehicle (VEH) was injected. After the first devaluation test, mice were re-trained with the VR schedule for one day. Then the second devaluation test was conducted with VEH or CNO injection in a cross-over design. **(B)** A representative image of hM4Di-mCherry expression in the ACC. **(C)** Effects of ACC inhibition on

lever pressing in the devaluation test. **(D)** Within-subject changes in the devaluation index of **(C)**. **(E)** Mice received injection with AAV-CaMKII-hM3Dq-mCherry in the ACC. The first devaluation test was performed on the following the 4th VI training. After the first test, mice were re-trained with the VI schedule for one additional day, after which the second devaluation test was performed. Same as **(A)**, drug injections were conducted in a cross-over design. **(F)** Representative image of hM3Dq-mCherry expression in the ACC. **(G)** Effects of the ACC activation on lever pressing in the devaluation tests. **(H)** Within-subject changes in the devaluation index of **(G)**. **(I)** Mice received injection with AAV-CaMKII-hM4Di-mCherry in the LOFC. The experimental protocol was same as **(A)**, except for the AAV injection sites. **(J)** Representative image of hM4Di-mCherry expression in the LOFC. **(K)** Effects of the LOFC inhibition on lever pressing in the devaluation tests. **(L)** Within-subject changes in the devaluation index of **(K)**. *$P < 0.05$; **$P < 0.01$.

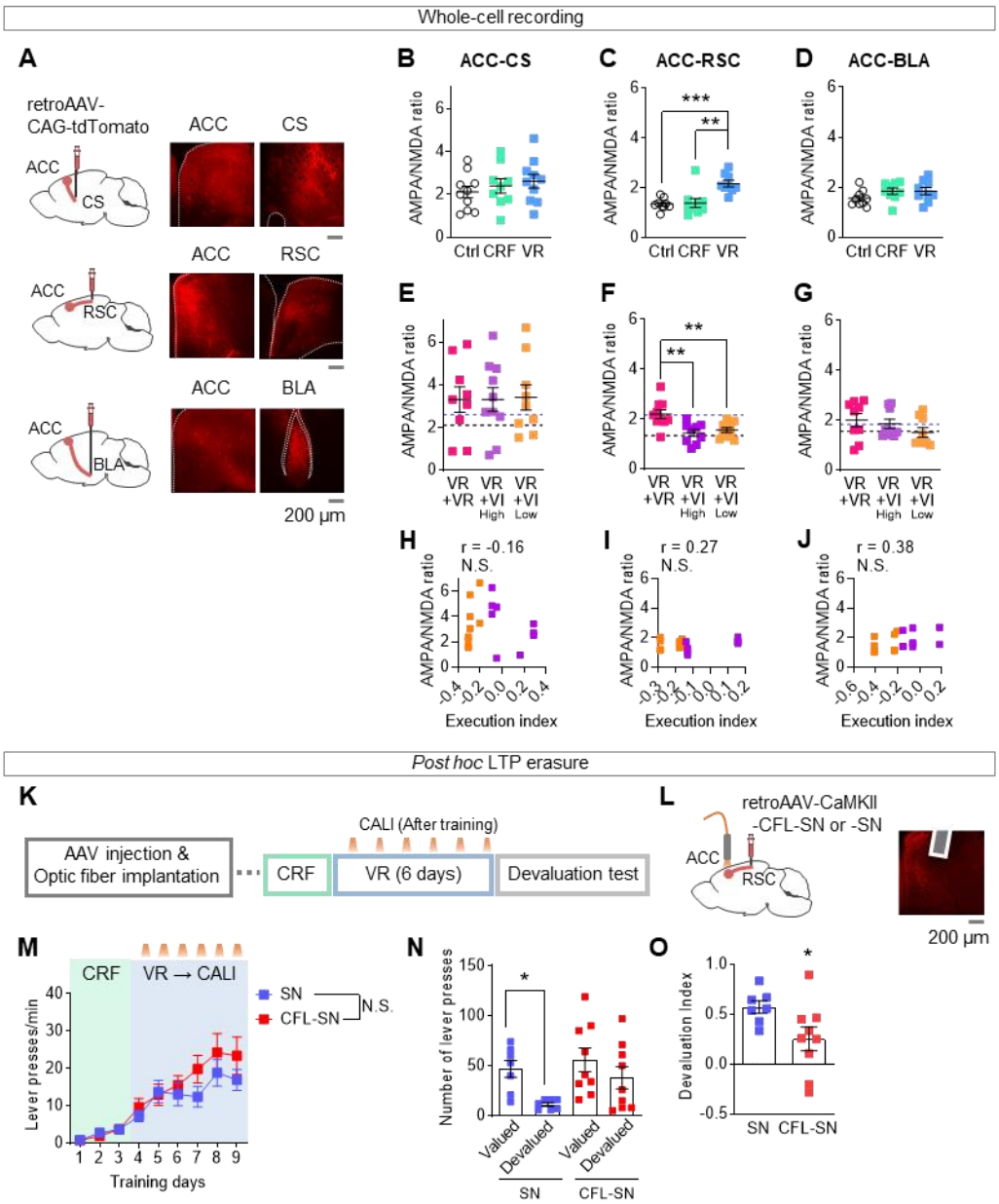


Figure 4

LTP in the RSC-projecting ACC neurons during VR training is necessary for maintenance of goal-directed decision-making

(A) Representative images of retrograde labeling with retroAAV-CAG-tdTomato. The retro AAV was injected in the central striatum (CS), retrosplenial cortex (RSC) and basolateral amygdala (BLA). For all three groups, tdTomato-positive neurons were detected in the ACC. **(B–D)** Acute brain slices of ACC were prepared within 30 min after the 3rd CRF and 6th VR training. Control (Ctrl) mice received food restriction but were not trained. AMPA/NMDA ratios were recorded from CS-projecting ACC

layer 5 neurons **(B)**, RSC-projecting ACC layer 5 neurons **(C)** and BLA-projecting ACC layer 5 neurons **(D)**. **(E–G)** Acute brain slices of the ACC were prepared within 30 min after the 4th VI training. The VR+VR group was introduced as a control matched for total training days. AMPA/NMDA ratios were recorded from CS-projecting ACC layer 5 neurons **(E)**, RSC-projecting ACC layer 5 neurons **(F)** and BLA-projecting ACC layer 5 neurons **(G)**. Black and blue dashed lines show the mean value of Ctrl and VR groups in **(B–D)** respectively. **(H–J)** Correlation between execution index and AMPA/NMDA ratios in the VR+VI$_{High}$ and VR+VI$_{Low}$ groups. **(K)** Schematic diagram of VR training with CALI. Mice received CALI (illumination with 1–2 mW amber light for 1 min) within 5-min after each VR training session ended (days 4–9). **(L)** Representative image of CFL-SN expression and optic fiber implantation in the ACC. retroAAV-CaMKII-CFL-SN or retroAAV-CaMKII-SN were injected into the RSC and optic fibers implanted into the ACC. **(M)** Lever press rate during training with CALI. **(N)** Effects of CALI in RSC-projecting ACC neurons on lever pressing in the devaluation tests. **(O)** Within-subject changes in the devaluation index of **(N)**. *P < 0.05; **P < 0.01; ***P < 0.001; N.S., not significant.

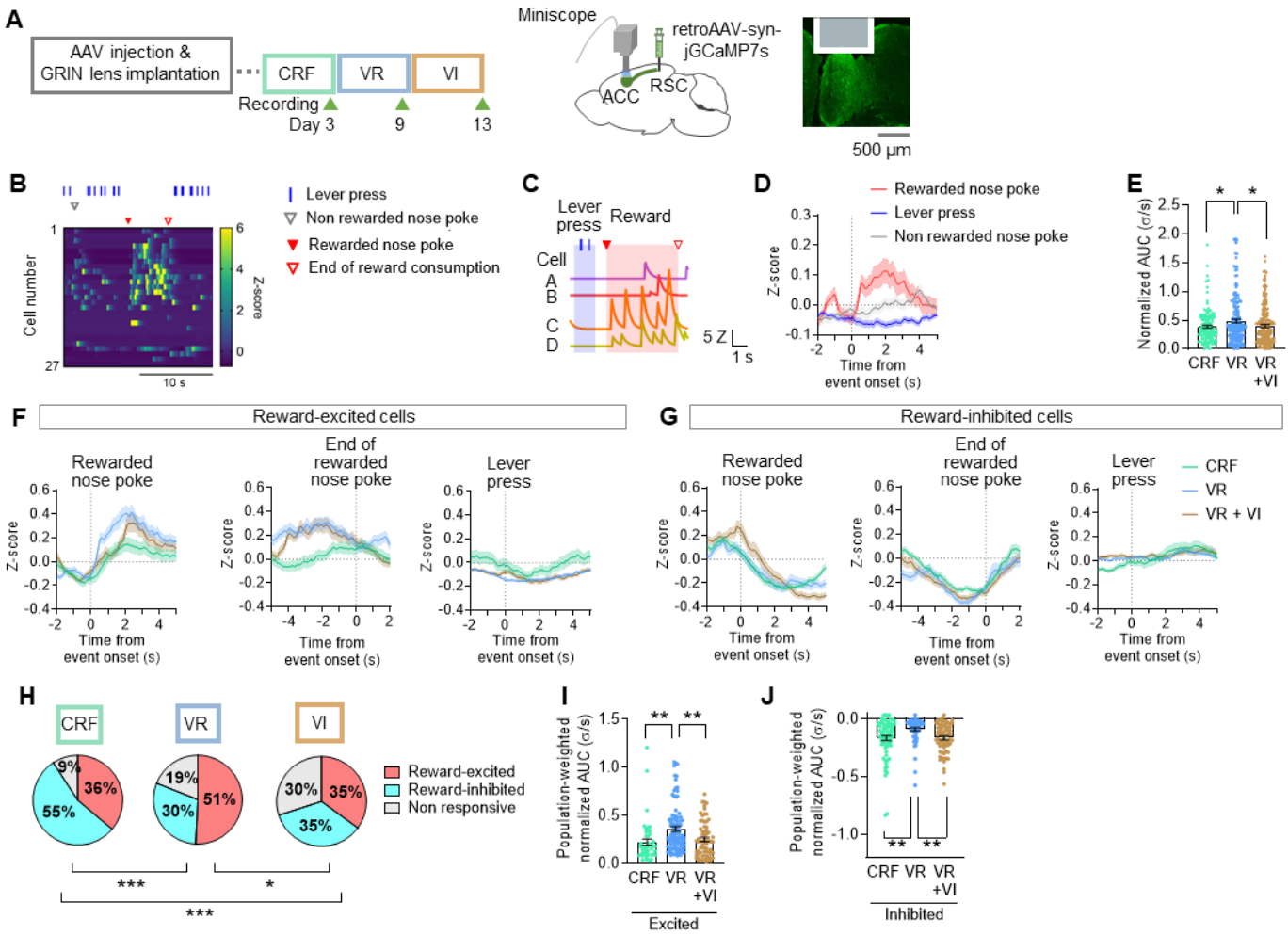
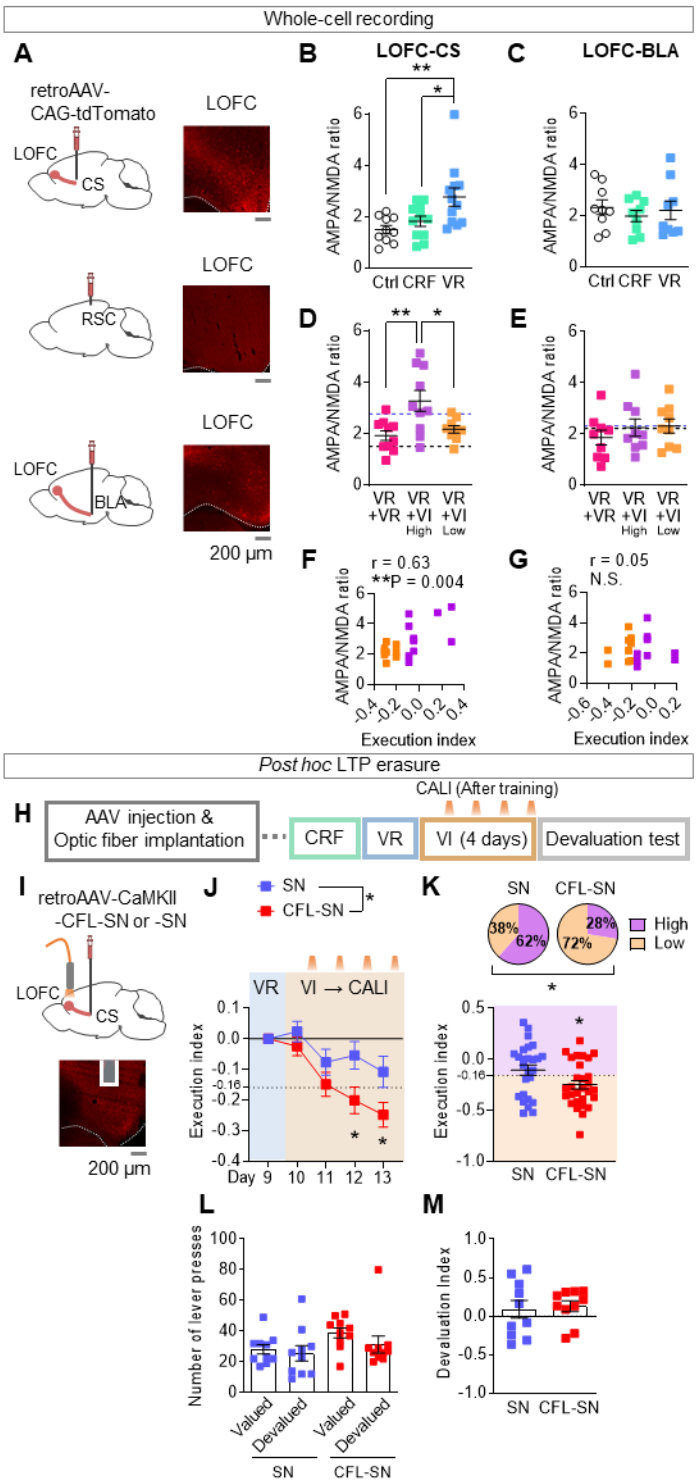


Figure 5

Reward-related activation of RSC-projecting ACC neurons is potentiated during goal-directed action

(A) Schematic diagram of *in vivo* Ca^{2+} recording and a representative image of jGCaMP7s expression and GRIN lens implantation. retroAAV-syn-jGCaMP7s was injected into the RSC and GRIN lens was implanted into the ACC. Recordings were performed in the last CRF (day 3), VR (day 9) and VI (day 13) training sessions. **(B)** Representative heat map of Ca^{2+} transients around the first, second (left) and tenth (right) rewarded nose pokes during VR training (day 9). A subset of RSC-projecting ACC neurons were activated during reward consumption. **(C)** The red and blue highlighted zones show the time windows for calculation of reward and lever press-related neural response, respectively (Method). **(D)** The average activity of all recorded cells during lever pressing, rewarded nose pokes and non-rewarded nose pokes during VR training. **(E)** The AUC of reward-related responses normalized by the duration of events (From the start to the end of rewarded nose pokes). Each dot

represents each cell's average of normalized AUCs for reward events during training. **(F,G)** The average activity of reward-excited **(F)** and -inhibited **(G)** cells at the start and end of rewarded nose pokes and lever pressing during each training stage. **(H)** Fractions of reward-excited and reward-inhibited cells at each training stage. **(I,J)** For reward-excited and reward-inhibited cells, the population-weighted normalized AUC were calculated by multiplying normalized AUCs of positive **(I)** or negative **(J)** reward-related responses by the fraction of the reward-excited and reward-inhibited cells. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.



1091 **Figure 6**

1092 *LTP in the CS-projecting LOFC neurons during VI training determined the execution of habitual*

1093 *responses*

1094 **(A)** Representative images for retrograde labeling of LOFC neurons with retroAAV-CAG-tdTomato.

1095 The retro AAV was injected in the central striatum (CS), retrosplenial cortex (RSC) and basolateral

amygdala (BLA). Each image was taken from the same mouse shown in Fig. 4A. tdTomato-positive neurons were not detected in the LOFC, when the retroAAV vector was injected in the RSC. **(B,C)** Acute brain slices of LOFC were prepared within 30 min after the 3rd CRF and 6th VR training. Control (Ctrl) mice underwent food restriction but were not trained. AMPA/NMDA ratios were recorded from CS-projecting LOFC layer 5 neurons **(B)** and BLA-projecting LOFC layer 5 neurons **(C)**. **(D,E)** Acute brain slices of the LOFC were prepared within 30 min after the 4th VI training. The VR+VR group was introduced as a control matched for total training days. AMPA/NMDA ratios were recorded from the CS-projecting LOFC layer 5 neurons **(D)** and BLA-projecting LOFC layer 5 neurons **(E)**. Black and blue dashed lines show the mean value of the Ctrl and VR groups in **(B,C)** respectively. **(F,G)** Correlation between the execution index and AMPA/NMDA ratios in the VR+VI$_{High}$ and VR+VI$_{Low}$ groups. **(H)** Schematic diagram of the two-step training protocol with CALI. Mice received CALI (illumination with 1–2 mW amber light for 1 min) within 5-min after each VI training (day 10–13). **(I)** Representative image of CFL-SN expression and optic fiber implantation in the LOFC. The retroAAV-CaMKII-CFL-SN or retroAAV-CaMKII-SN was injected into the CS and optic fibers were implanted into the LOFC. **(J)** Execution index during VI training with CALI. **(K)** Execution index on day 13 and the fraction of VR+VI$_{High}$ and VR+VI$_{Low}$ mice in each group. **(L)** Effects of CALI in the CS-projecting LOFC neurons on lever pressing in the devaluation test. **(M)** Within-subject changes in the devaluation index of **(L)**. *$P < 0.05$; **$P < 0.01$; N.S., not significant.

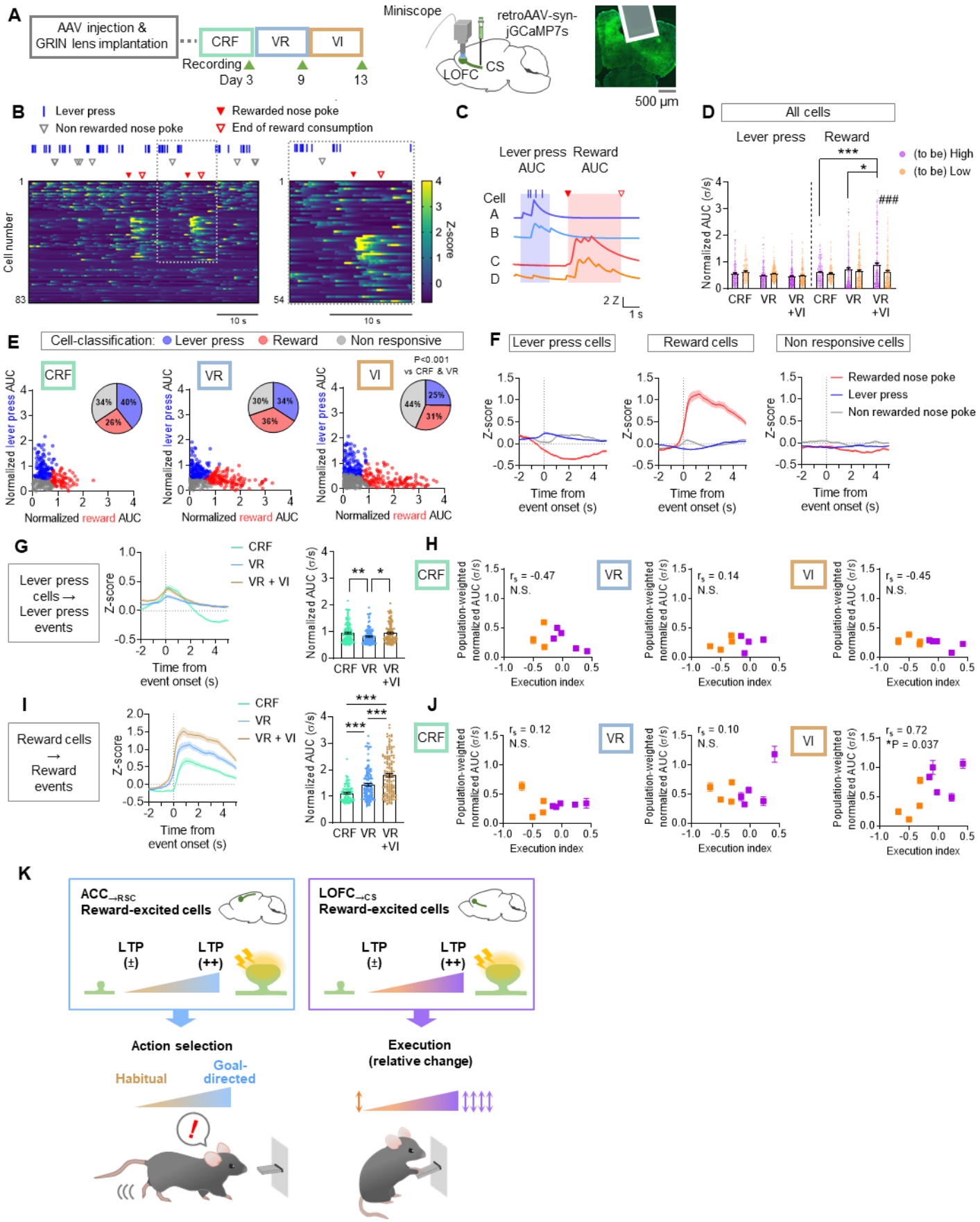

A
AAV injection & GRIN lens implantation
CRF
VR
VI
Recording
Day 3
9
13
Miniscope
retroAAV-syn-jGCaMP7s
LOFC
CS
500 µm
B
Lever press
Non rewarded nose poke
Rewarded nose poke
End of reward consumption
Cell number
Z-score
10 s
C
Lever press AUC
Reward AUC
Cell
A
B
C
D
2 Z
1 s
D
All cells
Lever press
Reward
(to be) High
(to be) Low
Normalized AUC (σ/s)
CRF
VR
VR +VI
E
Cell-classification: Lever press Reward Non responsive
P<0.001 vs CRF & VR
Normalized lever press AUC
Normalized reward AUC
F
Lever press cells
Reward cells
Non responsive cells
Rewarded nose poke
Lever press
Non rewarded nose poke
Z-score
Time from event onset (s)
G
Lever press cells → Lever press events
CRF
VR
VR + VI
Z-score
Time from event onset (s)
Normalized AUC (σ/s)
H
Population-weighted normalized AUC (σ/s)
Execution index
r_s = -0.47
N.S.
r_s = 0.14
N.S.
r_s = -0.45
N.S.
I
Reward cells → Reward events
J
r_s = 0.12
N.S.
r_s = 0.10
N.S.
r_s = 0.72
*P = 0.037
K
ACC→RSC Reward-excited cells
LOFC→CS Reward-excited cells
LTP (±)
LTP (++)
Action selection
Habitual
Goal-directed
Execution (relative change)

Figure 7

Reward-related activity of CS-projecting LOFC neurons is positively correlated with performance changes during habit formation

(A) Schematic diagram of *in vivo* Ca^{2+} recording and representative image of jGCaMP7s expression and GRIN lens implantation. The retroAAV-syn-jGCaMP7s was injected into the CS and GRIN lens was implanted into the LOFC. Recordings were performed in the last CRF (day 3), VR (day 9) and VI (day 13) training sessions. **(B)** Representative heat map of Ca^{2+} transients during VR training (day 9). The right panel shows the enlarged version of the dotted area on the left. **(C)** Representative trace of Ca^{2+} transients around lever pressing and reward events. Cells A and B were activated during lever presses, but not at reward consumption. In contrast, cells C and D showed higher activity during reward consumption than at lever presses. The red and blue highlighted zones show the time windows for calculation of reward and lever press-related neural response, respectively (Method). **(D)** The AUC of event-related responses were normalized by the duration of lever press bouts, or reward consumption. Each dot represents each cell's average of normalized AUCs for lever press or reward events during training. **(E)** The XY-planes represent lever press and reward-related normalized AUCs for each cell. Cells were classified into three subgroups according to the magnitudes and preference of their responses toward lever press and reward events (Method). The pie charts represent fractions of each subgroup. **(F)** Average activity of lever press cells, reward cells and non-responsive cells during each type of event during VR training. **(G)** Average activity and normalized AUCs of lever press cells to lever press events during each training stage. **(H)** Correlation between the execution index and population-weighted AUCs of the lever press cells to lever press events during each training stage. **(I)** Average activity and normalized AUCs of the reward cells to reward events during each training stage. **(J)** Correlation between the execution index and population-weighted AUCs of the reward cells to reward events during each training stage. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; N.S., not significant. For **(D)**, $^{\#}P < 0.05$ at (to be) High vs. (to be) Low. **(K)** Dynamic regulation of cortical region-specific synaptic potentiation orthogonally controls habitual behavior. LTP in RSC-projecting

ACC neurons is necessary for the predominance of goal-directed strategy for action selection, and as it weakened, behavior becomes habitual. Conversely, LTP in CS-projecting LOFC neurons serves to maintain the execution of the habitual behavior at a comparable level to its previous, goal-directed execution.