Dissociable roles of prefrontal plasticity in decision making strategy and execution of habitual behavior

Nozomi Asaoka*, Diane Pagano and Yasunori Hayashi*

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

Running title: **Dissociable cortical plasticity in habit regulation**

* 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)
0009-0000-5763-4811 (DP)
0000-0002-7560-3004 (YH)

Abstract

Habits are essential for sustaining adaptive behaviors but can also lead to maladaptive behaviors such as compulsive and addictive disorders. They emerge through a shift in decision-making from motivation-driven strategy to an automatic one. During this, the level of habit execution, such as frequency and duration, is superficially maintained despite a decline in motivational drive, raising the question of how the amount of execution is maintained even when decision-making strategies undergo substantial changes. By developing a unique paradigm capable of inducing habit formation within a defined time window in male mice, we found that shift in decision-making and amount of habit execution are independently controlled by plasticity in distinct cortical pathways. Erasure of each plasticity selectively altered the decision-making strategy or the execution level without affecting the other. These findings reveal a dual regulatory model for habit, providing new insights into the neurocircuit mechanisms underlying both adaptive and maladaptive habits.

Introduction

Habits play a crucial role in maintaining beneficial but effortful behaviors, while maladaptive habits cause refractory, over-engaged behaviors shown in compulsive disorders and behavioral addictions[1–5]. Elucidating the regulatory mechanisms of habitual behavior is therefore essential for understanding both adaptive and maladaptive habits[6–8]. Habit formation has been conceptualized as a shift in decision-making—from a flexible, motivation-dependent goal-directed strategy to a more rigid, automatic one[2,5,7]. While this framework captures qualitative differences in behavior selection, it overlooks the quantitative dimension of habitual execution, such as duration and frequency of the behavior, that is critical for evaluating the real-world utility of adaptive habits and the pathological severity of maladaptive ones. Indeed, during the transition, despite the decline in motivational drive, the level of execution is superficially maintained; however, the mechanism that allows the execution level to be preserved despite significant changes in decision-making strategies remains unclear.

This motivated us to propose that the assessment of habitual behavior should incorporate not only the nature of the underlying decision-making strategy but also an additional dimension: regulation of execution level. While decision-making and execution levels in goal-directed behavior are regulated by evaluations of cost and benefit[9,10], habitual actions, at least at the level of decision-making, proceed automatically without such value-based evaluations. Therefore, a distinct mode of control is likely to govern execution levels in habitual behavior. Based on this premise, we hypothesize the existence of a neural mechanism that stores and maintains the execution level of behavior independently of the mechanisms that govern the transition from goal-directed to habitual strategies.

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 execution level in the same cohort of animals. Our analyses revealed that the transition in decision-making strategy and execution level are indeed separately controlled. To analyze the circuit mechanism, we combined this task with *ex vivo* slice recordings, *in vivo* Ca^{2+} imaging and a recently-

61 developed optogenetic method to erase long-term potentiation (LTP). We found that predominance
62 of a goal-directed strategy and execution level are orthogonally represented as potentiation of synaptic
63 transmission in different prefrontal cortical regions. By targeting region specific plastic changes, we
64 were able to modify the decision-making strategy without affecting execution level or *vice versa*,
65 supporting the notion that decision strategy and execution level are distinct processes that involve
66 different circuits. Our findings on this dual regulatory model extend the current view of habit
67 regulation and offer new insights for understanding and treating maladaptive habits in disorders such
68 as obsessive-compulsive disorder and addiction.
69

Results

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

Previous studies have demonstrated that training with ratio-based and interval-based operant conditioning tasks have contrasting outcomes on animal behaviors[3,5,11]. With ratio-based tasks, where rewards are delivered after a certain number of lever presses, animals acquire goal-directed behavior. In contrast, with interval-based tasks, rewards are delivered on the first lever press after a specific amount of time has passed since the prior reward[12,13]. The number and frequency of lever pressing are less contingent on reward deliveries than in ratio-based tasks and the timing of reward deliveries is less affected by subjects' lever pressing effort. Owing to these factors, interval-based task designs make it difficult for subjects to understand the association between behavioral effort and reward acquisition. Instead, subjects focus on the behavior that previously led to reward delivery; resulting in the predominance of a habitual over goal-directed strategies[3,14].

To date, studies have used these two protocols separately to monitor the acquisition of goal-directed and habitual behaviors[3,5]. However, this approach makes identifying the timing of the transition from goal-directed to habitual behaviors difficult, and thus does not allow for the study of the neural mechanisms underlying this transition. Moreover, due to such unclear timing of habit formation, the quantitative changes in action execution accompanied by habit formation could not be adequately analyzed. To overcome these issues, we performed ratio- and interval-based tasks in a sequential fashion in the same cohort of mice and evaluated within-individual transition of decision-making strategy together with that of action execution (Fig. 1A). 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. Mice were first trained with a 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 a reward port. 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 variable time interval from the last reward (ranging from 30 to 90 s, average 60 s). When switched to the VI task, mice slightly but significantly decreased the frequency of lever pressing, compared with animals that continued the VR task (VR+VR group) (Fig. 1B).

To confirm the VR+VI task design led to a transition from a goal-directed to habitual strategy, we conducted two tests: a devaluation test and an omission test[15–17]. In the devaluation test, mice were given free access to either sucrose solution (devalued condition), or regular rodent chow (valued condition) for 30 minutes. They were then subjected to a 5-min lever-press task during which the sucrose reward was not delivered, without signaling this to the mice (Fig. 1C). When mice had already consumed sucrose solution, the sucrose reward is considered “devalued”. In contrast, when mice were pre-fed with rodent chow, the reward remains “valued”. On the following day, the same mice were tested in the opposite condition. When mice employ a goal-directed strategy, they 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, when they employ a habitual strategy, the number of lever presses is comparable between the valued and devalued conditions, because the reward value does not affect the decision. Because the absolute number of lever presses during the devaluation test is heavily influenced by the lever press rate during the preceding training, we defined the devaluation index for each mouse, which is the difference in lever pressing between the two conditions, normalized by the total number of lever presses (see Methods). This index gives a positive value when subjects rely on a goal-directed strategy and approaches zero when they employ a habitual strategy, irrespective of absolute lever press number[18]. 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 (1st test; Fig. 1D–F), confirming that mice employed a goal-directed strategy[5,17]. However, after an additional 4 days of VI training, mice pressed the lever a comparable number of times in both the valued and devalued conditions. This led to devaluation indices close to zero, indicating that mice had changed their strategy to a habitual one with VI training. This was not simply the result of prolonged training, because mice that underwent 13 days of VR training (VR+VR) continued to exhibit a reduction in the number of lever presses in the devalued condition (2nd test; Fig. 1D & F). The effects of devaluation in the VR+VI mice were comparable to those of mice trained for 13 days with the VI paradigm (VI+VI mice. Fig. S1A–C).

In contrast to lever pressing, significant decreases in the number of nose pokes to the reward delivery port were observed in the devalued condition for all groups, except for the first test of the VI+VI group, which exhibited a nonsignificant trend toward reduction. These results suggest that their motivation towards sucrose rewards was reduced irrespective of the training conditions (Fig. S1D–G). In addition, the sucrose/chow intake balance during free access did not differ between conditions (Fig. S1H). Therefore, 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 habitual decision strategy, which is not motivated by the current value of sucrose rewards. Consequently, these results confirm that VR→VI switching, but not extended VR+VR training leads to the formation of habitual behavior.

Habit formation was also assessed in the omission test[17]. In the task conducted on the day after the completion of VR+VI training, sucrose solution was delivered automatically every 20 seconds when the lever was not pressed, whereas lever pressing triggered a delay in reward delivery, thereby degrading the association between lever pressing and sucrose delivery (Fig. S1I). Mice utilising a goal-directed strategy diminish the number of lever presses more quickly than those adopting a habitual strategy[17]. Consistent with this idea, the number of lever presses decreased slower in the VR+VI group, compared with the VR+VR group (Fig. S1J & K).

145 Taken together, these data indicate that implementing a VI training period, following VR training
146 (VR+VI task) converts an established goal-directed behavior into habitual behavior, thereby
147 providing a versatile paradigm to allow for the detailed analysis of the regulatory mechanisms
148 underlying newly formed habitual behaviors.
149

150 **Execution level of habitual behavior is controlled independently from the decision-making**
151 **strategy**

152 Our VR+VI task data revealed there was considerable variation within-subjects in lever pressing
153 changes associated with the transition from a goal-directed to habitual decision strategy (Fig. 1G–O).
154 We hypothesized that this variation may prove valuable for identifying the neuronal mechanisms that
155 determine execution level. To this end, we calculated an execution index for each mouse by
156 normalizing the difference in the frequency of lever presses between days 9 (final VR session) and
157 13 (final VI session), by the total of the frequencies in the two sessions (Fig. 1G, Methods). The
158 execution index varied among mice trained in the VR+VI task from positive to negative with a median
159 value of -0.16 (max: 0.44, min: -0.58, standard deviation (SD): 0.21). In contrast, mice trained with
160 the VR+VR protocol showed significantly higher execution indices with a median of 0.11 (max: 0.71,
161 min: -0.09, SD: 0.18), indicating that the majority of mice maintained or increased their lever pressing
162 rate after 4 days of additional VR training (Fig. 1H).

163 Next, for further analysis, the VR+VI group was operationally subdivided into two groups;
164 $VR+VI_{High}$ and $VR+VI_{Low}$, based on execution indices falling above or below the median value (-
165 0.16) (Fig. 1H). The lever press rate on day 9 was comparable between the $VR+VI_{High}$ and $VR+VI_{Low}$
166 groups and did not affect this classification (Fig. 1I–K). Likewise, the number of lever press bouts
167 (see Methods for definition) was comparable on day 9 between the $VR+VI_{High}$ and $VR+VI_{Low}$ groups,
168 however, on day 13, it was significantly decreased in the $VR+VI_{Low}$ group (Fig. 1L). These data
169 indicate that there was no significant difference in behavior between the $VR+VI_{High}$ and $VR+VI_{Low}$

170 groups during goal-directed behavior, prior to the VR→VI switching. As for other behavioral metrics,
171 there was no significant difference in the frequencies of rewards acquisitions nor nose pokes between
172 the VR+VI_{High} and VR+VI_{Low} groups (Fig. 1M & N). Although there were slight decreases in nose
173 poke frequencies over the course of the training period, the index of nose poke frequency calculated
174 using the same method as the execution index showed no significant correlation with the execution
175 index (Fig. 1O). These observations indicate that differences in execution index between the
176 VR+VI_{High} and VR+VI_{Low} groups resulted from habit formation, rather than reflecting pre-existing
177 variability during goal-directed behavior nor other variations in reward-consuming behaviors.
178 To test if the strength of habitual strategy correlates with the variation observed in execution indices,
179 we compared the results of the devaluation and omission tests between the VR+VI_{High} and VR+VI_{Low}
180 groups. In the devaluation test, both groups displayed a significant inhibitory effect before VR→VI
181 switching, but not after VR→VI switching (Fig. 1P & Q). Both before and after switching, there was
182 no significant correlation between the devaluation index and execution index (Fig. 1R). Also in the
183 omission test, the decay in lever pressing was comparable across the VR+VI_{High} and VR+VI_{Low}
184 groups, both of which exhibited a slower decay than the VR+VR group (Fig. S1L–N). These data are
185 consistent with the idea that the strength of the habitual strategy does not correlate with execution
186 level.
187 We next tested if the variation in execution index arises from other behavioral factors by using a T-
188 maze spatial discrimination and reversal learning test and marble burying test, which can measure
189 learning ability, cognitive flexibility and compulsive-like behavior, respectively, after completion of
190 the VR+VI task (Fig. S1O & P). In all experiments, there was no significant difference between the
191 VR+VI_{High} and VR+VI_{Low} groups, suggesting that the differences in execution indices were
192 determined by the daily training, rather than the individual's general characteristics of cognition or
193 decision-making.

194 Collectively, these results suggest that the VR→VI switching protocol induces a transition from a
195 goal-directed strategy to a habitual strategy. However, the bias toward habitual strategy was not
196 correlated with behavioral execution level, indicating that they are likely governed by independent
197 mechanisms.
198
199 **Cortical region-specific potentiation of synaptic responses in specific task stages**
200 Since repeated training was required for both behavioral learning and habit formation, we
201 hypothesized that training induces plastic changes in synaptic transmission. To search for evidence
202 that plastic changes were responsible for deciding strategy and execution level, excitatory
203 postsynaptic currents (EPSCs) were recorded *ex vivo* from pyramidal neurons in the prefrontal
204 cortical regions. We focused on anterior cingulate, lateral and medial orbitofrontal cortices (ACC,
205 LOFC and MOFC), as they have been implicated in reinforcement learning and decision-making[10,19].
206 Mice were sacrificed after CRF (day 3), VR (day 9) and VR+VR or VR+VI (day 13) training, after
207 which AMPA/NMDA ratios were recorded from pyramidal neurons in layer 2/3 (L2/3) and layer 5
208 (L5), to assess the induction of synaptic potentiation of excitatory synaptic transmission[20,21] (Fig. 2A–
209 L).
210 Following CRF training, there was a specific increase in AMPA/NMDA ratio in MOFC L2/3
211 neurons, but not in other regions or layers (Fig. 2A–F). However, in the group that underwent
212 subsequent VR training, this increase was diminished. Instead, there were increases in the
213 AMPA/NMDA ratios in both L5 neurons of the ACC and LOFC (Fig. 2B & D). This suggests that as
214 goal-directed behavior is progressively acquired, synaptic potentiation begins in MOFC L2/3 neurons,
215 but transitions to encompass primarily the ACC and LOFC L5 neurons.
216 Following VR→VI switching (Fig. 2G–L), the AMPA/NMDA ratio in ACC L5 neurons returned to
217 control levels in both the VR+VI_{High} and VR+VI_{Low} groups, whereas it remained elevated in the
218 VR+VR group (Fig. 2G & H). However, the VR+VI_{High} group showed significantly higher

219 AMPA/NMDA values in the LOFC L5 neurons, but not L2/3, compared with both the VR+VR and
220 VR+VI_{Low} groups (Fig. 2I & J). In the MOFC, there were no significant differences in AMPA/NMDA
221 ratios among groups (Fig. 2K & L). Finally, when comparing individual task performance, we
222 observed a significant correlation between the AMPA/NMDA ratio of LOFC L5 neurons and
223 execution index (Fig. 2M–R).

224 To confirm the findings observed with evoked EPSCs, we recorded Sr^{2+}-induced AMPAR-mediated
225 asynchronous EPSCs (aEPSCs) (Fig. S2A–F). The results were consistent with the evoked EPSC data,
226 except for a high aEPSC amplitude observed from LOFC L5 neurons in the VR+VR group (Fig. S2F),
227 which contradict the evoked EPSCs that were comparable to control levels. In the late stages of LTP,
228 NMDA currents are potentiated following potentiation of AMPA currents, resulting in a reset of the
229 AMPA/NMDA ratio[21]. Therefore, it is possible that NMDAR inputs to those LOFC L5 neurons had
230 also been potentiated by prolonged VR+VR training, potentially leading to an overall decrease in
231 AMPA/NMDA ratios, compared with the VR group.

232 We also tested for changes in presynaptic release probability using the paired-pulse ratio (PPR; Fig.
233 S2G–R). There were no significant differences in the PPR in all regions tested, except for a significant
234 decrease in ACC L5 pyramidal neurons of the VR group, suggesting that presynaptic release
235 probability is also increased.

236 Together, our observations suggest that region-specific plastic changes in the cortex potentially
237 underlie transitions between goal-directed and habitual strategies, as well as modulating execution
238 levels. Acquisition of a goal-directed behavior potentiated AMPAR-mediated synaptic transmission
239 in the ACC and LOFC L5 neurons. In the ACC L5 neurons, these synaptic inputs were subsequently
240 depotentiated by habit formation. In contrast, execution correlated with potentiation of LOFC L5
241 neurons, but not other regions tested.

242

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

245 To test if the increased excitability of ACC neurons was required for the predominance of goal-
246 directed over a habitual strategy, we employed a chemogenetic approach. The inhibitory receptor
247 hM4Di was expressed in ACC and mice were subsequently trained in the VR task for acquisition of
248 a goal-directed behavior. Next, mice were administered with either vehicle or clozapine-*N*-oxide
249 (CNO) to inhibit ACC neurons prior to the devaluation test (Fig. 3A–D). After the first devaluation
250 test, mice were re-trained with the VR protocol for a day, then underwent a second devaluation test
251 with the reverse drug condition. Lever pressing in mice that received the vehicle was significantly
252 decreased in the devalued condition, however the CNO treated group showed no significant changes
253 between the valued and devalued conditions, indicative of a predominantly habitual strategy.
254 Conversely, when the activity of ACC neurons was increased with hM3Dq after the transition to
255 habitual behavior by the VR+VI task, devaluation induced a significant reduction in lever pressing in
256 the devaluation test, indicating that habitual behavior could be reverted to a goal-directed strategy via
257 artificial ACC modulation (Fig. 3E–H). Additionally, similar effects of ACC activation on the
258 devaluation test were consistently observed in mice trained only with the VI task for 6 days,
259 suggesting that the contribution of ACC in promoting a goal-directed strategy remains irrespective of
260 the training paradigm used (Fig. S3A–C). In all the three experiments, two-way repeated measures
261 ANOVA revealed no significant main effect of CNO administration, suggesting that the effects of
262 chemogenetic manipulations were mainly derived from neural inhibition/activation rather than side
263 effects of CNO (Table S1).

264 Interestingly, although VR training induced synaptic potentiation in the LOFC (Fig. 2J & P),
265 chemogenetic inhibition did not affect the devaluation-induced response, unlike ACC inhibition (Fig.
266 3I–L). Moreover, chemogenetic manipulation of LOFC during the omission test didn't significantly

267 alter lever pressing when comparing the vehicle- and CNO-treated hM4Di or hM3Dq groups (Fig.
268 S3D–I).
269 During the chemogenetic experiments, CNO administration did not affect preference or satiety for
270 sucrose rewards in the devaluation tests, as no significant changes were detected in sucrose/chow
271 intake balance, or reduction in nose poking at the reward delivery port in the devalued condition (Fig.
272 S3J–M).
273 Together, these observations indicate that the suppression of ACC activity can convert behavior from
274 a goal-directed to habitual strategy, whereas activation drives goal-directed behavior. By contrast, the
275 excitability of LOFC neurons does not lead to shifts in decision strategy.
276
277 **Projection-specific plasticity of ACC neurons in specific training stages**
278 ACC neurons are a heterogenous population in terms of both gene expression and projection
279 patterns[22,23]. To determine whether training-induced synaptic potentiation occurs in a specific
280 subclass of ACC neurons, which target a particular region, or if it is more global encompassing the
281 entire ACC, we injected retrograde AAV encoding tdTomato into the major targets of ACC: the central
282 striatum (CS), retrosplenial cortex (RSC), and the basolateral amygdala (BLA) to label each subclass
283 of ACC projection neuron[24–26] (Fig. 4A). Following this, training-induced synaptic potentiation was
284 assessed in tdTomato-positive L5 pyramidal neurons by recording the AMPA/NMDA ratio (Fig. 4B–
285 J). A significant increase in the AMPA/NMDA ratio after VR training was detected only in the RSC-
286 projecting L5 neurons (Fig. 4C). After VR→VI switching, the AMPA/NMDA ratio in this class of
287 neuron returned to control levels, in both the VR+VI_{High} and VR+VI_{Low} groups, whereas the VR+VR
288 group remained elevated (Fig. 4F). We observed no significant correlations between the
289 AMPA/NMDA ratio and execution indices for any subclass of projection neuron tested (Fig. 4H–J).
290 These results indicate that synaptic potentiation in ACC-RSC projection neurons plays an essential

291 role in the acquisition of goal-directed behavior and that its de-potentiation is required for habit
292 formation.
293
294 **Optical erasure of synaptic potentiation in RSC-projecting ACC neurons facilitates the**
295 **transition to habit**
296 To further test this idea, we employed the recently-developed optogenetic tool, cofilin-SuperNova
297 (CFL-SN) to erase long-term potentiation (LTP). Cofilin is an actin-binding protein that accumulates
298 in dendritic spines and stabilizes filamentous actin after the induction of LTP[27]. Inactivation of cofilin
299 by chromophore-assisted light inactivation (CALI) using a photosensitizer protein SuperNova (SN)
300 erases LTP *post hoc,* without affecting basal synaptic transmission or the stability of pre-existing
301 synapses[28]. CFL-SN was expressed in RSC-projecting ACC neurons to determine if *post hoc* LTP
302 erasure would shift the decision strategy from a goal-directed to habitual strategy. To this end,
303 retrograde AAV was used to express CFL-SN, or SN as a control in the RSC-projecting ACC neurons
304 bilaterally (Fig. 4K & L). After each VR session (days 4–9), light was delivered to the ACC through
305 optical fibers to erase training-induced LTP. After a 6-day VR training protocol, a devaluation test
306 was conducted to assess decision strategy. During the VR training, both the CFL-SN and SN groups
307 increased the frequency of lever presses (Fig. 4M), however, in the devaluation test the CFL-SN group
308 showed a comparable number of lever presses in the valued and devalued conditions and a
309 significantly smaller devaluation index than the SN group, indicating that their behavior became
310 habitual (Fig. 4N & O). Since, the two groups had a comparable level of lever pressing during training,
311 this indicates that the acquisition of the lever pressing behavior itself does not require LTP in RSC-
312 projecting ACC neurons (Fig. 4M). Additionally, CALI did not impact the preference and satiety for
313 sucrose rewards in the devaluation processes, suggesting that the evaluation of reward value remains
314 intact (Fig. S4).

315 Overall, the results suggest that training-induced excitatory synaptic potentiation in the RSC-
316 projecting ACC neurons is required for the maintenance of a goal-directed strategy, but if lost, a
317 habitual strategy becomes predominant. This is consistent with the observation that using VI training
318 to establish habitual behavior, reduced the AMPA/NMDA ratio in this class of neuron (Fig. 4B–J).
319

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

321 Our data highlight the importance of the potentiation of synaptic transmission in a specific class of
322 ACC neuron projecting to RSC, in determining transitions in decision-making strategy. Given their
323 importance and to better understand how these neurons influence behavior, we carried out Ca^{2+}-
324 imaging in freely moving animals using a head-mounted miniature fluorescent microscope[29].
325 RSC-projecting ACC neurons were targeted by injecting a retrograde AAV vector encoding
326 jGCaMP7s in the RSC (Fig. 5A). A subset of these neurons exhibited burst-like activity during
327 rewarded nose pokes, whereas neurons showed minimal activity during lever pressing or non-
328 rewarded nose pokes (Fig. 5B–D). The mean activity of neurons during rewarded nose pokes was
329 significantly larger during VR training, but reduced following VR→VI switching (Fig. 5E). We
330 classified RSC-projecting ACC neurons into reward-excited and -inhibited cells (Fig. 5F–H;
331 Methods). Significant changes in the fraction of each subclass were observed during the transition
332 from goal-directed to habitual behavior (Fig. 5H). The proportion of reward-excited cells (Fig. 5H)
333 as well as their activity (Fig. 5I) were increased by VR training and decreased following VR→VI
334 switching (Fig. 5I), consistent with our *ex vivo* slice electrophysiology data. In contrast, the proportion
335 of reward-inhibited neurons (Fig. 5H) as well as their reward-evoked inhibition (Fig. 5J) decreased
336 during VR training, but reverted following VR→VI switching. The increase in reward-related
337 excitation followed by goal-directed lever pressing and its reversion during habitual behavior, align
338 with the view that a goal-directed strategy relies on current reward value in decision-making, whereas
339 a habitual strategy makes decisions independent of the value of a reward. There was no significant

correlation between changes in activity of these neurons and execution index (Fig. S5). 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 VR+VI_{Low} animals, thereby specifically correlating with execution index after VR→VI switching (Fig. 2J, P). We therefore further explored how LOFC neurons may regulate changes in action execution accompanied by habit formation.

To achieve this, the same retrograde labeling strategy was used to identify specific subclasses of LOFC neurons projecting to CS, RSC, and BLA. We found LOFC projections to both CS and BLA, but not to the RSC (Fig. 6A). The training 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 excitatory transmission of CS-projecting LOFC neurons determines the execution of habitual behavior (Fig. 6F & G).

Training-induced persistent synaptic potentiation of the CS-projecting LOFC neurons maintains 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 erased LTP in these neurons after each VI session (days 10–13; Fig. 6H & I). After this period, the average execution index and proportion of VR+VI_{High} mice in the CFL-SN

365 group was significantly lower than the control (SN) group (Fig. 6J & K). Additionally, the effects of
366 LTP erasure by CALI on the expression of habitual behavior was assessed by devaluation and
367 omission tests. In the devaluation test, there was no significant difference in the number of lever
368 presses between the valued and devalued conditions in both the SN and CFL-SN groups (Fig. 6L).
369 This resulted in the comparable devaluation indices between the SN and CFL-SN groups (Fig. 6M).
370 In addition, CALI in CS-projecting LOFC neurons had no significant effect on motivation toward
371 sucrose reward (Fig. S6A & B). Likewise, in the omission test (Fig. S6C & D), CALI had no
372 significant effect on the decay of lever press rate quantified as reduction in relative lever pressing rate
373 (Fig. S6C & D).
374 Next, to test if the increased excitability of CS-projecting LOFC neurons during habitual task
375 performance is sufficient for facilitation of action execution, we performed chemogenetic activation
376 of CS-projecting LOFC neurons during VI session (Fig. S6E–L). After completion of the 13-day
377 VR+VI task, mice received additional two VI sessions with CNO or vehicle injection. CNO
378 administration induced a slight but significant increase in lever pressing frequency (Fig. S6F), without
379 affecting frequency of reward acquisitions or nose pokes (Fig. S6G & H). The same manipulation of
380 CS-projecting LOFC neurons had no significant effect on lever-press responses or motivation toward
381 sucrose reward in devaluation test (Fig. S6I–L).
382 These results indicate that training-induced potentiation in CS-projecting LOFC neurons during
383 habit formation, specifically determines changes in execution level without affecting the transition
384 from goal-directed to habitual strategies.
385
386 **Reward-related activity in the CS-projecting LOFC neurons after habit formation correlates**
387 **with the execution index**
388 To confirm this idea, activity of CS-projecting LOFC neurons during training were recorded using
389 Ca^{2+}-imaging (Fig. 7A). At each step of the training, significant responses were associated with both

390 lever pressing and rewarded-nose pokes (Fig. 7B–D). Responses to lever pressing and rewarded-nose
391 pokes arose in two non-overlapping populations of neurons, which were classified as lever press and
392 reward neurons respectively (Fig. 7E & F; Methods). This contrasts with RSC-projecting ACC
393 neurons which responded only to the reward, but not lever pressing (Fig. 5A–G). In addition, the
394 reward-responding CS-projecting LOFC neurons showed a strong response with slow decay
395 following the onset of rewarded nose poke (Fig. 7B, C). This also highlights distinct characteristics
396 from reward-excited RSC-projecting ACC neurons, which frequently exhibit multiple, brief activities
397 during reward consumption (Fig. 5B–D). The overall activity of neurons during lever pressing was
398 unchanged throughout the training session (Fig. 7D & E). In contrast, the activity during the
399 rewarded-nose poke increased specifically in the VR+VI_{High} group, but not in the VR+VI_{Low} group
400 (Fig. 7D). These neurons were mutually inhibited; the lever press neurons were inhibited during the
401 firing of reward neurons and *vice versa* (Fig. 7F). Following a transition from goal-directed to habitual
402 behavior, significant changes in the fraction of both classes of neurons were observed (Fig. 7E) but
403 no correlation was observed between the fractions of cells and execution index of each mouse (Fig.
404 S7A, B). Lever press cell activity was relatively stable at different phases of training, with the
405 exception of a small but statistically significant decrease during the VR training (Fig. 7E & G). There
406 was no significant change in the lever press-related normalized AUC of the lever press cells and
407 correlation with execution index of each mouse (Fig. 7H, Fig. S7C). In contrast, the amplitude of the
408 reward-related Ca^{2+} transients progressively increased across the training stages (Fig. 7E & I).
409 Additionally, at the VI training stage, but not VR stage, the reward-related response had a positive
410 correlation with the execution index of each mouse (Fig. 7J), suggestive of a causal relationship
411 between reward-related activity of CS-projecting LOFC cells and execution index of habitual
412 behavior. Comparing the reward-related activity of each mouse across training stages revealed no
413 consistent enhancement between the VR and VI training stages, suggesting that the positive
414 correlation with the execution index cannot be attributed solely to an increase in the number of

415 training sessions (Fig. S7D). Taken together with the whole-cell recording and LTP erasure data (Fig.
416 6), reward-related activation of CS-projecting LOFC neurons plays a pivotal role in sustaining a
417 consistent level of execution during habit formation.
418

419 **Discussion**

420 Identifying the regulatory mechanisms that operate during the transition from goal-directed to
421 habitual behavior has proven difficult, primarily due to the lack of an appropriate behavior model
422 where habit formation can be induced at a defined time window. Our newly designed task, which
423 sequentially combines VR and VI tasks, overcomes this limitation by including clearly separable
424 goal-directed and habit formation phases. Using this task, we demonstrated that habitual behavior
425 should be considered from two mechanistic dimensions, the decision-making strategy and execution
426 level, that are independently controlled. Supporting this, further analyses revealed that decision-
427 making strategy and execution level are orthogonally represented in the brain. At the circuit level, the
428 emergence of goal-directed behavior is mediated by the potentiation of synaptic transmission in RSC-
429 projecting ACC neurons. Then the transition from a goal-directed to habitual strategy is accompanied
430 by a depotentiation of transmission in those same neurons. In contrast, execution level is determined
431 by the persistence/absence of synaptic potentiation in CS-projecting LOFC neurons. We confirmed
432 the necessity and sufficiency of these circuits by chemogenetically manipulating the activity, and
433 additionally by preventing training associated plastic changes. Furthermore, *in vivo* Ca^{2+}-imaging
434 revealed changes in neuronal firing patterns, which together demonstrate that distinct populations of
435 prefrontal cortical neurons govern predominance among decision-making strategies and execution
436 level of the habitual behaviors independently.

437 The ACC is required for monitoring and prediction of an action-outcome relationship[30,31], by
438 representing various reward-related information, such as reward value and cost of action[32–34].
439 Consistent with this, we found that chemogenetic manipulation of ACC activity bidirectionally
440 regulated goal-directed/habitual decision strategy. Though recent studies have demonstrated that
441 striatum-projecting ACC neurons are essential for cost-benefit assessment, but not goal-directed
442 decision-making[35], and BLA-projecting neurons represent the innate value of reward and aversive
443 signals[36,37], the mechanisms governing goal-directed decision-making have remained elusive. In this

444 regard, this is the first demonstration that plasticity in RSC-projecting ACC neurons can dictate goal-
445 directed strategies for decision-making. Since both the ACC and RSC are crucial for learning and
446 memory of associative information, which is transferred from the hippocampus[28,38], it is conceivable
447 that information of action-outcome association for goal-directed decision-making, which is
448 transiently encoded in the hippocampus at earlier stages of training[39], is transferred to the RSC-
449 projecting ACC neurons.

450 Erasure of the VR training-induced LTP in the RSC-projecting ACC neurons did not impair the
451 acquisition of lever pressing behavior, while facilitating habit formation. Consistent with this
452 observation, previous behavioral studies in rodents have consistently shown that disruption of brain
453 regions implicated in goal-directed strategy, including the dorsomedial striatum, preserves
454 reinforcement learning, yet facilitates habit formation[5,40]. These findings suggest that when the goal-
455 directed strategy is compromised, animals instead recruit reinforcement learning processes biased
456 toward habitual behavior. Supporting this, computational studies have proposed two parallel forms
457 of reinforcement learning, model-based and model-free (or value-free) learning corresponding to
458 goal-directed and habitual strategies respectively, that develop concurrently during behavior
459 acquisition[2,41,42]. Moreover, human studies have shown that habit formation is blocked while model-
460 based learning is predominant, suggesting inhibitory mechanisms of the goal-directed strategy on the
461 habitual strategy[43]. Accordingly, the intact reinforcement learning in the LTP erasure group is most
462 consistent with preferential disruption of model-based learning system.

463 In contrast to the ACC, although a number of studies implicate the OFC in reward-related decision-
464 making, its functional significance in goal-directed decision-making has been controversial; both
465 activation and inhibition of the OFC disrupt flexibility of a behavior and shift it to a habitual one[5,44].
466 In addition, other prior studies have shown the OFC has no, or a limited contribution to behavior
467 flexibility and defining decision strategies[45,46]. While possible causes for this inconsistency, such as

468 heterogeneity of the OFC subdivisions and testing protocols, have been postulated[5,44,45,46], a specific
469 role for the OFC in goal-directed strategies is still to be fully elucidated.

470 Despite the functional differences between RSC-projecting ACC neurons and CS-projecting LOFC
471 neurons, Ca^{2+}-imaging revealed that both neurons represent reward (outcome)-related information.
472 Considering that *post hoc* LTP erasure was sufficient to affect the decision strategy and execution
473 level, training-induced learning and updating of outcome-related association, rather than innate
474 representation of reward value, may be critical for the effects of both neurons. As conceptualized by
475 computational studies[2,41,46], rewarded experience updates the model for stimulus-response-outcome
476 associations when using a model-based learning strategy. Hence, outcome-related information
477 encoded in the RSC-projecting ACC neurons is likely to be utilized for updating the model of
478 stimulus-response-outcome associations.

479 In contrast, outcome-related information encoded in the CS-projecting LOFC neurons is less likely
480 to directly modulate the habitual strategy, since this system utilizes stimulus-response associations
481 rather than action-outcome associations. In line with this, comparative studies of ACC and OFC
482 highlight their dissociative roles in encoding outcome value; the OFC encodes outcome-related
483 information during Pavlovian (cue-outcome) learning rather than during the instrumental (action-
484 outcome) learning, while the opposite contribution is found in the ACC[47,48]. Pavlovian learning is
485 known to facilitates execution of operant behavior through Pavlovian-instrumental transfer,
486 particularly when the response is habitual[49,50]. Although our task was not explicitly designed to induce
487 Pavlovian conditioning, interoceptive cues during tasks such as hunger could be associated with
488 reward presentation. This raises the possibility that cue-outcome encoding in the CS-projecting LOFC
489 neurons facilitates execution level of habitual behavior. In addition, it has been reported that the OFC
490 is critical for contextual knowledge of outcome expectation, which hierarchically regulates
491 performance of already acquired behaviors[51,52]. This function of the OFC has been mainly discussed
492 in relation to flexible adaptation of behaviors according to context discrimination, where subjects

493 learn distinct associations in a context-specific manner[5,52]. By contrast, when a behavior is stably
494 rewarded in a specific context, this contextual stability has been proposed to facilitate execution of a
495 habitual behavior, possibly through serving the contextual information as a promoter[8,14,53]. Taken
496 together, LTP in CS-projecting LOFC neurons could contribute to the association between rewarded
497 experience and its preceding cue or context, thereby facilitating the execution of lever pressing habit
498 without affecting the stimulus-response associations.

499 The regulation of goal-directed and habitual behaviors by cortico-striatal pathways is well
500 established[8,10]. Accumulating evidence supports dichotomous control mechanisms, such as
501 "dorsomedial versus dorsolateral striatum" and "associative versus sensorimotor circuits"[54,55].
502 However, recent histological analyses revealed distinct projections to striatal regions from each
503 cortical region[56,67] suggesting that these histologically separated pathways have distinct functions.
504 Consistently, it has repeatedly been reported that LOFC neurons send inputs to a subregion of the
505 striatum ranging from the dorsomedial to the central striatum. Dense inputs are observed in the central
506 subregion[57–60]. This unique projection pattern of the LOFC supports the present finding that CS-
507 projecting LOFC neurons are implicated in habit execution rather than decision-making strategy.

508 Accumulating evidence from animal and clinical research has shown that hyperactivity in the OFC-
509 striatum pathway plays a pivotal role in the pathology of compulsivity[61–63]. Abnormally strong habit
510 is considered as a possible cause of excessively repeated execution of certain behavior(s) shown in
511 patients with compulsivity[64,65]. Symptom-related, cue-induced hyperactivity in the OFC and
512 increased functional connectivity between the OFC and striatum have been reported in patients with
513 compulsive disorders[66–69]. While present study focuses on habitual behavior in healthy mice, the
514 concept of LOFC-derived individuality in habit execution may be applicable to the pathological
515 habitual behaviors such as obsessive-compulsive disorder, addiction and eating disorders.

516 In conclusion, the present study provides novel insights into neural mechanisms governing
517 behavioral control during the goal-directed to habitual transition. The spatiotemporally distinct online

518 encoding of outcome-related information in the ACC and LOFC was required for blocking habit
519 formation and maintenance of action execution, respectively (Fig. 7K). Our findings highlight the
520 ACC-RSC pathway as a critical determinant of decision-making strategy, and further, reveal an
521 unexpected role of LOFC-CS pathway on controlling habit execution.

Acknowledgments

We thank Drs. Dai Watanabe, Tomoyuki Furuyashiki, Chika Nishimura, 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 and JP25K02556 from the MEXT, Japan, AMED Brain/MINDS 2.0 JP24wm0625507, Uehara Memorial Foundation, Kobayashi Foundation, Takeda Science Foundation and Astellas Foundation for Research on Metabolic Disorders (to N.A.) and Grants-in-Aid for Scientific Research JP18H05434, JP22H04981 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 (to Y.H.).

Author contributions

N.A. and Y.H. designed the research. N.A. and D.P. performed the experiments and collected data. N.A. and Y.H. conceived and conducted behavioral and neural analyses. N.A. and Y.H. wrote the manuscript.

Competing interests

The authors declare no conflicts of interest.

Supplemental information

Document S1. Figure S1–7.

Table S1. Excel file containing statistical information.

546 **Materials and Methods**

547 *Animals*

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

555

556 *Operant training*

557 Operant training was carried out using an operant chamber (Med Associates, St Albans, VT)
558 controlled by MED-PC software (version IV and V; Med Associates). Each session began with
559 turning on the house light and lever presentation and ended when mice received the maximum number
560 of reinforcers, or when 60 min had elapsed, unless otherwise described. For magazine training (2 or
561 3 (for Ca^{2+} recordings) sessions), food-restricted mice were placed in an operant chamber without
562 lever presentation, after which 16 reinforcers (10 μL of 20% sucrose solution) were presented at 60-
563 s intervals on average. For habituation to the lever, mice received three continuous reinforcement
564 (CRF) sessions, where each lever press was rewarded. The max number of reinforcers increased
565 across sessions (5, 15 and 30 reinforcers). Following CRF sessions, mice were trained with the VR
566 or VI protocols. During VR training, an average of 10 (VR10) or 20 (VR20) presses were required
567 for each reinforcer, whereas for VI training the reward was delivered on first press after an average
568 interval time of either 30 (VI30) or 60 (VI60) sec following delivery of the previous reward. The sets
569 of variables were fixed, but their order was changed across sessions. Mice that underwent the VR10
570 or VI30 schedule for the first two VR or VI sessions; thereafter were trained with the VR20 or VI60

571 schedule. When mice were unable to obtain 10 reinforcers on the 4th VR20 or VI60 session, they were
572 excluded from further experiments.
573 For the two-step paradigm, mice received three CRF, two VR10 and four VR20 sessions, which
574 were followed by four VI60 sessions. The transition from VR20 to VI60 schedules were seamlessly
575 carried out; without any cue or context changes. The changes in lever press rates between the last VR
576 (fourth VR20) session and last VI60 session (fourth VI60 session) were indexed as follows;
577 Execution index = [(press rate in VI session) – (press rate in the VR session)]/ [(press rate in the VI
578 session) + (press rate in the VR session)].
579 For Fig. 7C, press rate in each VI60 session, was used instead of the last VI60 session.
580 All lever presses, reward deliveries and nose pokes during training were automatically recorded by
581 the MED-PC software. The end of reward consumption was manually detected as the time at which
582 mice start to move their head away from the reward port. For analyzing lever press bouts, presses
583 with an inter-event interval of < 0.5 s were considered as one bout.
584
585 *Devaluation test*
586 After the last training session, devaluation tests were conducted over two consecutive days. Mice
587 were allowed free access to food chew (the valued day) or a 20% sucrose solution (the devalued day)
588 for 30 min. The order of the valued and devalued days was randomized. After the free access session,
589 mice were placed in the operant chamber for 5 min where the lever was presented as usual, but never
590 reinforced. To normalize the baseline variability of lever pressing activity among individuals, the
591 difference in lever pressing numbers between the valued and devalued conditions for each mouse was
592 presented as the devaluation index. The devaluation index was calculated as follows:
593 Devaluation index = [(lever presses on the valued day) – (lever presses on the devalued day)]/ [(lever
594 presses on the valued day) + (lever presses on the devalued day)].
595 One mouse was excluded from experiments shown in Fig. S6J–M because the pellet intake during

596 the free access session was too low to detect.

597

598 *Omission test*

599 After the last training session, an omission test was conducted. During the 30-min omission session
600 the lever was presented as usual. When the lever was not pressed for 20 s, a reward (20% sucrose
601 solution) was provided at the reward port, however when the lever was pressed a 20-second cooldown
602 timer was reset.

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

606

607 *Spatial discrimination learning and reversal learning*

608 After completion of the two-step operant training, T-maze spatial discrimination training was
609 performed as previously described[70]. Food restriction was continued throughout all behavioral
610 experiments. Mice were habituated with the T-maze apparatus (one start arm (30 × 10 cm), two goal
611 arms (30 × 10 cm) and 30-cm-high surrounding walls) for 7 days and trained for spatial discrimination
612 for 12 days (5 trials/day). The rewarded arm (rewarded with 100 μL of sweetened milk) was randomly
613 chosen and fixed during the 12-day training sessions. For reversal learning, the rewarded arm was
614 reversed, and mice underwent 10 free-choice trials per day for 4 days.

615

616 *Marble burying test*

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

621

622 *Stereotaxic surgery*

623 Mice were anesthetized with isoflurane (1–2%) and fixed on a stereotaxic frame (Narishige, Tokyo,
624 Japan). Adeno-associated virus (AAV) vector (0.3 μL/side) was injected with a stainless-steel cannula
625 (30 G) at the following coordinates, unless otherwise described; ACC (AP 1.2 mm, ML ±0.5 mm, DV
626 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
627 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
628 mm). The injection site for the CS was selected based on previous reports[58–60].

629

630 *DREADD experiments*

631 For global manipulation of ACC/LOFC pyramidal neurons, the AAV vector expressing hM4Di
632 (AAV-CaMKIIα-hM4Di-mCherry; Addgene #50477) or hM3Dq (AAV-CaMKIIα-hM3Dq-mCherry;
633 Addgene #50476) was bilaterally injected into the ACC or LOFC. For manipulation of the CS-
634 projecting LOFC neurons, the AAV vector expressing hM3Dq (AAV-DIO-hSyn-hM3Dq-mCherry;
635 Addgene #44361) and retrograde AAV vector expressing Cre (AAV.CamKII 0.4.Cre.SV40; Addgene
636 #105558) was bilaterally injected into the LOFC and CS, respectively. Mice were allowed to recover
637 for at least one week prior to starting the operant training; so that at least 4-weeks had passed before
638 CNO was injected.

639 For the devaluation tests, CNO (0.5 mg/kg; Cayman Chemical, MI, USA) or its vehicle (1% DMSO
640 in saline) was injected prior to starting the free access session. The dose of CNO was chosen based
641 on previous studies[5, 72]. For the cross-over design, mice were re-trained with one VR20 session after
642 the first set of devaluation tests. The second set of devaluation tests were performed under the
643 opposite drug condition. The order of the valued and devalued conditions was the same as in the first
644 set. For manipulation during the training and omission test, CNO or its vehicle were injected 20 min
645 before starting the session.

646

647 *In vivo Ca^{2+} imaging*

648 Mice received unilateral injections of retroAAV-syn-jGCaMP7s-WPRE (Addgene #104487) in either
649 the RSC or CS. To obtain a sufficient number of jGCaMP7s-positive cells, AAV was injected at two
650 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 &
651 0.2 mm, ML ±1.7 mm, DV 4.0 mm). Next a GRIN lens (1-mm diameter, 4-mm length; Inscopix, CA,
652 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
653 ±1.7 mm, DV 1.0 mm), after which mice were singly housed until experiments concluded.

654 At least 3 weeks post injection, a baseplate for the UCLA Miniscope[29] (version 4.4 purchased from
655 OpenEphys) was fixed over the GRIN lens with dental cement. Lens implantation and baseplate
656 mounting were performed with slight modifications to the method described at http://Miniscope.org.
657 Operant training began at least 4 weeks after AAV injection. To enable mice with a Miniscope to
658 enter the reward port, a wide custom-made nose poke hole was used. Mice underwent the two-step
659 operant task with three magazine training sessions. For habituation to the Miniscope, mice were
660 trained with a dummy scope after the second magazine training session. On the 3rd CRF, 6th VR and
661 4th VI sessions (the last session of each training schedule), calcium transients during the session were
662 recorded at 30 Hz using the UCLA Miniscope DAQ software (version 1.10). Recordings of calcium
663 transients and animal behaviors were started and finished simultaneously with the operant training
664 sessions, by TTL signals from the MED-PC V software. Contrast and brightness of the videos were
665 adjusted using FFmpeg and semi-automatically processed with the MiniAn pipeline[71] for removing
666 sensor noise and background fluorescence, motion correction and source separation were performed
667 by a constrained non-negative matrix factorization (CNMF) algorithm. The spatial footprint and
668 calcium signal for each detected cell were manually verified to exclude false-positive detections.
669 While we maintained the same imaging fields across sessions, individual neurons were not
670 longitudinally tracked due to minor mechanical displacements of the imaging window and the focal

plane between recording days. Instead, all comparisons across training stages were conducted as population-level analyses.

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 – μ) / σ, 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 nose poked during this window, the event was excluded.

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, event-related negative AUC below the mean Ca^{2+} activities of pre-event periods, was divided by event length.

Non-rewarded nose poke: To avoid contamination of rewarded nose poke events, nose pokes that performed between lever presses that did not trigger reward presentation were classified as a non-rewarded nose poke. Based on this definition, non-rewarded nose pokes were analyzed only in the

696 VR and VI sessions. The time window was set to ±0.5 s around the nose poke, if further nose pokes
697 occurred within a 0.5 s interval, only the first one was included in analyses.
698
699 *Analysis of Ca^{2+} transients in RSC-projecting ACC neurons*
700 RSC-projecting ACC neurons were separated into three subclasses (reward-excited, reward-inhibited
701 and non-responsive cells).
702 Reward-excited cells: Neurons which met the following three criteria were classified as reward-
703 excited cells; (1) > 1 σ peak amplitude in at least 20% of rewarded nose poke events, (2) mean
704 amplitude during rewarded nose pokes was larger than that during pre-nose poke periods, (3) showing
705 multiple peaks during at least one rewarded nose poke event (Fig. 5D). Since, typical rewarded nose
706 pokes lasted 2-20 s, the criteria (2) and (3) were used to exclude a subset of neurons which showed
707 non-event-related episodic Ca^{2+} activity.
708 Reward-inhibited cells: Neurons that were not classified as reward-excited cells were further tested
709 for their decreases in Ca^{2+} activities during rewarded nose poke events. For each cell, the minimum
710 value of Ca^{2+} activity during each rewarded nose poke event was compared with the mean amplitude
711 during the corresponding pre-nose poke period using a paired *t*-test. When there was a significant
712 decrease in Ca^{2+} activity, the neuron was classified as a reward-inhibited cell.
713 For further statistical analyses, excitatory/inhibitory magnitude of each reward-excited/inhibited cell
714 was scored as a population-weighted response (Fig. 5I & J), which was calculated by normalized
715 AUC of rewarded nose pokes, multiplied by the fraction of reward-excited/inhibited cells in each
716 mouse[73].
717
718 *Classification of activity of CS-projecting LOFC neurons*
719 CS-projecting LOFC neurons were separated into three subclasses (reward, lever press and non-
720 responsive cells), as follows.

721 First, normalized AUCs of Ca^{2+} activities for reward and lever press events were calculated. Among
722 neurons recorded in the VR session, time-normalized AUCs for each event were Z-scored, so that
723 features of response selectivity could be compared. To this end, neurons recorded in the VR session
724 were clustered into 5 clusters by the K-means+ method (Python with scikit-learn package). The
725 number of clusters was determined by the elbow method. Then, to apply this classification to
726 recordings in the CRF and VI sessions, we made a logistic regression model for classification of
727 neurons using time-normalized AUCs datasets. For training, 75% of the datasets in the VR session
728 were used. The accuracy of the model tested with the remaining 25% of the datasets was 0.89.
729 Using the classification, neurons were separated into 5 groups: with high and moderate preference
730 for lever pressing, high and moderate preference for reward and non-responsive for both events. Since
731 the numbers of high preference neurons were small, high and moderate preference groups for each
732 event were combined. For each event-related neurons normalized AUC and population weighted
733 normalized AUC were calculated as described above.
734
735 *CALI experiments*
736 pAAV-CaMKIIα-CFL-SN and pAAV-CaMKIIα-SN were constructed by replacement of the hM4Di-
737 mCherry gene in the pAAV-CaMKIIα-hM4Di-mCherry (Addgene #50477) virus, to the CFL-SN and
738 SN genes, respectively. Retrograde AAV vectors for CFL-SN and SN expression were produced as
739 previously described[70].
740 Mice received bilateral injections of retroAAV-CaMKIIα-CFL-SN or retroAAV-CaMKIIα -SN in the
741 RSC or CS. Fiber optic cannulae, constructed from ceramic ferrules (1.25 mm diameter; CFLC270-
742 10; Thorlabs NJ, USA) and plastic optic fibers (250 μm diameter; CK-10; Mitsubishi Rayon, Tokyo,
743 Japan), were implanted in the ACC (AP 1.2 mm, ML ±0.5 mm, DV 1.0 mm, angled 10° relative to
744 the ML axis) or LOFC (AP 2.7 mm, ML ±1.7 mm, DV 2.5 mm). 595 nm laser light intensity from the
745 tip of the fiber optic cannula was adjusted to 1–2 mW. Operant training commenced at least 1-week

746 post-surgery, so that at least 3 weeks had elapsed since injection of virus.

747 For CALI in the RSC-projecting ACC neurons, mice received bilateral illumination of the ACC for
748 1 min, within 5 min after VR training had ended. For CALI in the CS-projecting LOFC neurons, mice
749 received bilateral illumination of the LOFC for 1 min, within 5 min after VI training had ended.

750

751 *Histology*

752 After completing all behavioral experiments, mice were deeply anesthetized with isoflurane and
753 transcardially perfused with PBS followed by 4% paraformaldehyde. Whole brain samples were
754 further fixed in 4% paraformaldehyde for 30 min at 4°C and incubated in 15% sucrose/PBS solution
755 overnight at 4°C. Coronal brain sections (30-μm thick) were prepared with a cryostat (Leica
756 CM1950/Leica CM1860; Leica Biosystems, Nußloch, Germany). Signals from fluorescent proteins
757 were visualized by a confocal microscopy system (FLUOVIEW FV1200, Olympus, Tokyo, Japan;
758 ECLIPSE Ti2-E, Nikon, Tokyo, Japan). For visualizing jGCaMP7s, sections were stained with rabbit
759 anti-GFP antibody (1:500; A11122, lot# 2659306; Thermo Fisher Scientific, MA, USA) followed by
760 AlexaFluor 594 conjugated secondary antibodies (1:200; Cell Signaling Technology, MA, USA).
761 Animals were excluded if post hoc histological analysis revealed viral expression or fiber or lens
762 placement outside the target region.

763

764 *Electrophysiology recordings*

765 Electrophysiology recordings were performed as previously described[70,74] with minor modifications.
766 For the control group, mice were food restricted for at least 2 weeks, but did not receive operant
767 training. In the projection-specific subpopulation groups, mice were injected with retroAAV-CAG-
768 tdTomato (Addgene #59462) at least 3 weeks before recording.

769 Within 30-min after the last operant training session, mice were deeply anesthetized with isoflurane
770 and decapitated. Coronal brain slices (200-μm thick) were prepared with a vibratome (VT1000S,

771 Leica Biosystems) filled with ice-cold cutting solution (in mM: 120 NMDG-Cl, 2.5 KCl, 26 $NaHCO_3$,
772 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
773 recovered in ACSF (composition in mM: 124 NaCl, 3 KCl, 26 $NaHCO_3$, 1 NaH_2PO_4, 2.4 $CaCl_2$, 1.2
774 $MgCl_2$, and 10 D-glucose, pH 7.3, oxygenated with 95%O_2/5% CO_2) at 32°C for at least 1 h before
775 recording. After recovery, individual slices were transferred to a recording chamber with continuous
776 perfusion of oxygenated ACSF, heated to 27 ± 1°C.

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

784 To record EPSCs, membrane potentials were held at -70 mV and a $GABA_A$ antagonist (20 μM
785 bicuculline) was continuously perfused. AMPA-mediated evoked EPSCs (eEPSCs) and mixed AMPA
786 and NMDA-mediated eEPSCs were evoked by stimulation at -70 mV and +40 mV, respectively.
787 NMDA-mediated eEPSC amplitude was determined as the average amplitude between 45 and 55 ms
788 post-stimulation. The average of 5 consecutive AMPA- and NMDA-mediated eEPSC measurements
789 was used for analyzing AMPA/NMDA ratios. AMPA receptor (AMPAR)-mediated eEPSCs were
790 normalized by NMDA receptor (NMDAR)-mediated eEPSCs. During paired pulse ratio experiments,
791 electrical stimuli were delivered with a 30-ms interval. The average of 5 consecutive PPRs was used
792 for analysis.

793 When recording asynchronous EPSCs, $CaCl_2$ in the ACSF was substituted with 6 mM $SrCl_2$.
794 Asynchronous events evoked by electrical stimulation were recorded 10 times from each neuron and
795 analyzed by Minianalysis software (SynaptoSoft, Decatur, GA) and Win EDR 3.7.3 (University of

796 Strathclyde). Events with amplitudes larger than 15 pA were used for analysis.

797

798 *Statistical analyses*

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

804

805 **Data availability**

806 The raw imaging and behavioral datasets are too large to share publicly. However, they are available
807 from the corresponding author upon reasonable request for research purposes. Source data are
808 provided with this paper.

809

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), 312-325. https://doi.org/10.1016/j.neuron.2013.09.007
3. Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. Philos Trans R Soc Lond B Biol Sci, 308(1135), 67–78. https://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 in cognitive sciences, 26(5), 371–387. https://doi.org/10.1016/j.tics.2022.02.002
5. Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nature communications, 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 psychology review, 9(3), 277–295. https://doi.org/10.1080/17437199.2013.876238
7. Wood, W., & Rünger, D. (2016). Psychology of Habit. Annual review of psychology, 67, 289–314. https://doi.org/10.1146/annurev-psych-122414-033417
8. Buabang, E. K., Donegan, K. R., Rafei, P., & Gillan, C. M. (2025). Leveraging cognitive neuroscience for making and breaking real-world habits. Trends in cognitive sciences, 29(1), 41–59. https://doi.org/10.1016/j.tics.2024.10.006
9. Courtin, J., Bitterman, Y., Müller, S., Hinz, J., Hagihara, K. M., Müller, C., & Lüthi, A. (2022). A neuronal mechanism for motivational control of behavior. Science (New York, N.Y.), 375(6576), eabg7277. https://doi.org/10.1126/science.abg7277
10. Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology, 35(1), 48–69. https://doi.org/10.1038/npp.2009.131
11. Yamada, K., & Toda, K. (2023). Habit formation viewed as structural change in the behavioral network. Communications biology, 6(1), 303. https://doi.org/10.1038/s42003-023-04500-2
12. Hilário, M. R., Clouse, E., Yin, H. H., & Costa, R. M. (2007). Endocannabinoid signaling is critical for habit formation. Frontiers in integrative neuroscience, 1, 6. https://doi.org/10.3389/neuro.07.006.2007
13. 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. Frontiers in integrative neuroscience, 4, 17. https://doi.org/10.3389/fnint.2010.00017
14. Bouton M. E. (2021). Context, attention, and the switch between habit and goal-direction in behavior. Learning & behavior, 49(4), 349–362. https://doi.org/10.3758/s13420-021-00488-z
15. 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. Behavioural brain research, 166(2), 189–196. https://doi.org/10.1016/j.bbr.2005.07.012

16. 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. Current biology: CB, 32(5), 1175–1188.e5. https://doi.org/10.1016/j.cub.2022.01.055

17. Rossi, M. A., & Yin, H. H. (2012). Methods for studying habitual behavior in mice. Current protocols in neuroscience, Chapter 8, Unit8.29–8.29. https://doi.org/10.1002/0471142301.ns0829s60

18. 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

19. Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nature neuroscience, 11(4), 389–397. https://doi.org/10.1038/nn2066

20. Takahashi, T., Svoboda, K., & Malinow, R. (2003). Experience strengthening transmission by driving AMPA receptors into synapses. Science (New York, N.Y.), 299(5612), 1585–1588. https://doi.org/10.1126/science.1079886

21. Watt, A. J., Sjöström, P. J., Häusser, M., Nelson, S. B., & Turrigiano, G. G. (2004). A proportional but slower NMDA potentiation follows AMPA potentiation in LTP. Nature neuroscience, 7(5), 518–524. https://doi.org/10.1038/nn1220

22. 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. Nature neuroscience, 19(12), 1733–1742. https://doi.org/10.1038/nn.4417

23. 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., … 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.2021.04.021

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

25. 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. Molecular pain, 18, 17448069221094529. https://doi.org/10.1177/17448069221094529

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

27. 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), 444–459. https://doi.org/10.1016/j.neuron.2014.03.021

28. 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

29. 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), 115–118. https://doi.org/10.1038/nature17955

30. 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.2023.08.022

31. Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J., & Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. Nature neuroscience, 9(7), 940–947. https://doi.org/10.1038/nn1724

32. 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.2020.10.013

33. Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., & Rushworth, M. F. (2006). Separate neural pathways process different decision costs. Nature neuroscience, 9(9), 1161–1168. https://doi.org/10.1038/nn1756

34. Monosov I. E. (2017). Anterior cingulate is a source of valence-specific information about value and uncertainty. Nature communications, 8(1), 134. https://doi.org/10.1038/s41467-017-00072-y

35. 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.2023.10.020

36. Ulloa Severino, F. P., Lawal, O. O., Sakers, K., Wang, S., Kim, N., Friedman, A. D., Johnson, S. A., Sriworarat, C., Hughes, R. H., Soderling, S. H., Kim, I. H., Yin, H. H., & Eroglu, C. (2023). Training-induced circuit-specific excitatory synaptogenesis in mice is required for effort control. Nature communications, 14(1), 5522. https://doi.org/10.1038/s41467-023-41078-z

37. Jhang, J., Lee, H., Kang, M. S., Lee, H. S., Park, H., & Han, J. H. (2018). Anterior cingulate cortex and its input to the basolateral amygdala control innate fear response. Nature communications, 9(1), 2744. https://doi.org/10.1038/s41467-018-05090-y

38. 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. Proceedings of the National Academy of Sciences of the United States of America, 116(17), 8576–8581. https://doi.org/10.1073/pnas.1818432116

39. Bradfield, L. A., Leung, B. K., Boldt, S., Liang, S., & Balleine, B. W. (2020). Goal-directed actions transiently depend on dorsal hippocampus. Nature neuroscience, 23(10), 1194–1197. https://doi.org/10.1038/s41593-020-0693-8

40. Giovanniello, J. R., Paredes, N., Wiener, A., Ramírez-Armenta, K., Oragwam, C., Uwadia, H. O., et al. (2025). A dual-pathway architecture for stress to disrupt agency and promote habit. Nature, 640(8059), 722-731. https://doi.org/10.1038/s41586-024-08580-w

41. 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

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

43. 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.

44. 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.

45. 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.

46. 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.

47. Camille, N., Tsuchida, A., & Fellows, L. K. (2011). Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. J Neurosci, 31(42), 15048-52. https://doi.org/10.1523/JNEUROSCI.3164-11.2011.

48. Luk, C. H., & Wallis, J. D. (2013). Choice coding in frontal cortex during stimulus-guided or action-guided decision-making. The Journal of neuroscience: the official journal of the Society for Neuroscience, 33(5), 1864–1871. https://doi.org/10.1523/JNEUROSCI.4920-12.2013

49. Wiltgen, B. J., Sinclair, C., Lane, C., Barrows, F., Molina, M., & Chabanon-Hicks, C. (2012). The effect of ratio and interval training on Pavlovian-instrumental transfer in mice. PLoS One, 7(10), e48227. https://doi.org/10.1371/journal.pone.0048227

50. Sebold, M., Schad, D. J., Nebe, S., Garbusow, M., Jünger, E., Kroemer, N. B., et al. (2016). Don't Think, Just Feel the Music: Individuals with Strong Pavlovian-to-Instrumental Transfer Effects Rely Less on Model-based Reinforcement Learning. J Cogn Neurosci, 28(7), 985-95. https://doi.org/10.1162/jocn_a_00945

51. 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

52. 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: CB, 34(23), 5532–5545.e3. https://doi.org/10.1016/j.cub.2024.10.049

53. 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

54. Malvaez, M., & Wassum, K. M. (2018). Regulation of habit formation in the dorsal striatum. Curr Opin Behav Sci, 20, 67-74. https://doi.org/10.1016/j.cobeha.2017.11.005

55. Vandaele, Y., Mahajan, N. R., Ottenheimer, D. J., Richard, J. M., Mysore, S. P., & Janak, P. H. (2019). Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. eLife, 8, e49536. https://doi.org/10.7554/eLife.49536

56. Foster, N. N., Barry, J., Korobkova, L., Garcia, L., Gao, L., Becerra, M., et al. (2021). The mouse cortico-basal ganglia-thalamic network. Nature, 598(7879), 188-194. https://doi.org/10.1038/s41586-021-03993-3

57. Hunnicutt, B. J., Jongbloets, B. C., Birdsong, W. T., Gertz, K. J., Zhong, H., & Mao, T. (2016). A comprehensive excitatory input map of the striatum reveals novel functional organization. eLife, 5, e19103. https://doi.org/10.7554/eLife.19103

58. Burguière, E., Monteiro, P., Feng, G., & Graybiel, A. M. (2013). Optogenetic stimulation of lateral orbitofronto-striatal pathway suppresses compulsive behaviors. Science, 340(6137), 1243-6. https://doi.org/10.1126/science.1232380

59. Corbit, V. L., Manning, E. E., Gittis, A. H., & Ahmari, S. E. (2019). Strengthened Inputs from Secondary Motor Cortex to Striatum in a Mouse Model of Compulsive Behavior. The Journal of neuroscience: the official journal of the Society for Neuroscience, 39(15), 2965–2975. https://doi.org/10.1523/JNEUROSCI.1728-18.2018

60. Hadjas, L. C., Schartner, M. M., Cand, J., Creed, M. C., Pascoli, V., Lüscher, C., & Simmler, L. D. (2020). Projection-specific deficits in synaptic transmission in adult Sapap3-knockout mice. Neuropsychopharmacology, 45(12), 2020-2029. https://doi.org/10.1038/s41386-020-0747-3

61. 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(6137), 1234–1239. https://doi.org/10.1126/science.1234733

62. 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: the journal of the European College of Neuropsychopharmacology, 25(11), 1898–1905. https://doi.org/10.1016/j.euroneuro.2015.08.017

63. 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(7736), 366–371. https://doi.org/10.1038/s41586-018-0789-4

64. 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

65. 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.2019.01.046

66. 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: official publication of the American College of Neuropsychopharmacology, 33(6), 1391–1401. https://doi.org/10.1038/sj.npp.1301513

67. 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. Journal of psychiatry & neuroscience: JPN, 33(5), 405–412.

68. 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

69. 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

70. 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. The Journal of neuroscience: the official journal of the Society for Neuroscience, 41(12), 2780–2794. https://doi.org/10.1523/JNEUROSCI.2121-20.2021

71. 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

72. Seabrook, L. T., Naef, L., Baimel, C., Judge, A. K., Kenney, T., Ellis, M., Tayyab, T., Armstrong, M., Qiao, M., Floresco, S. B., & Borgland, S. L. (2023). Disinhibition of the orbitofrontal cortex biases decision-making in obesity. Nature neuroscience, 26(1), 92–106. https://doi.org/10.1038/s41593-022-01210-6

73. 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/s41586-022-04954-0

74. 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. Translational psychiatry, 14(1), 27. https://doi.org/10.1038/s41398-024-02731-3

FIGURE LEGENDS

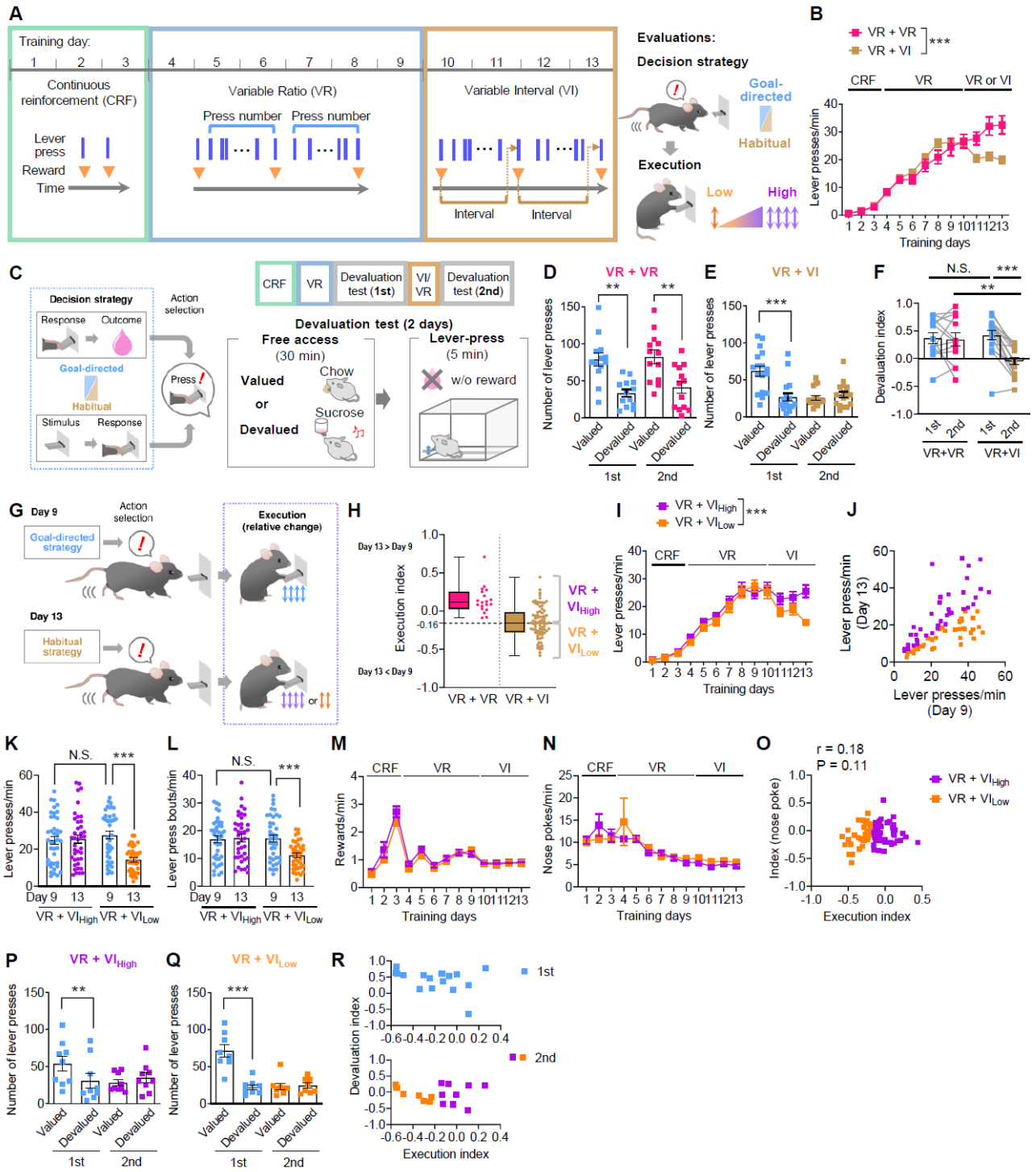


Figure 1

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

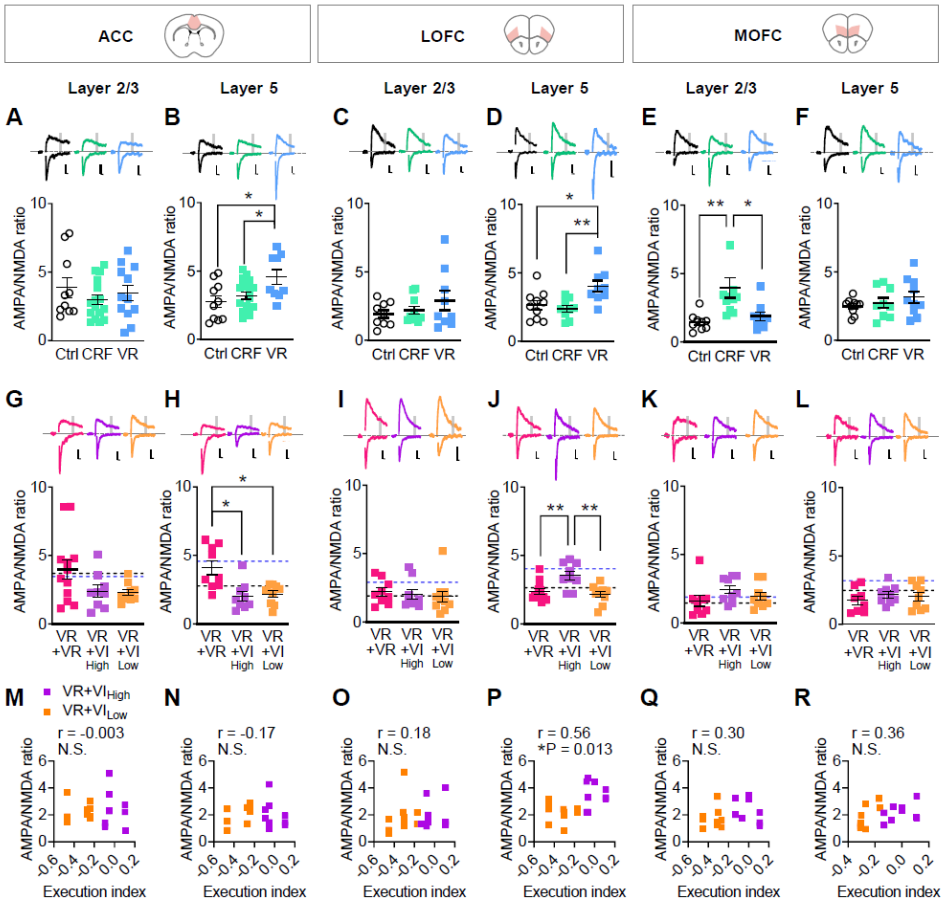
(A) Schematic diagram of the two-step operant training protocol. On day 10, the training schedule was seamlessly switched from a variable ratio (VR) to a variable interval (VI) task (VR+VI). Mice

without the VR→VI switching (VR+VR) are considered as the control group. **(B)** Transitions of lever pressing rate in the two-step training. VR+VR, N= 19; VR+VI, N= 79. **(C)** Schematic diagram of the devaluation test for evaluation of decision system. 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. (D–F) VR+VR, N=13; VR+VI, N=17. **(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 lever pressing bout in the $VR+VI_{Low}$ group. **(M)** Reward presentation rates in the $VR+VI_{High}$ and $VR+VI_{Low}$ groups. **(N)** Nose poke rates in the $VR+VI_{High}$ and $VR+VI_{Low}$ groups. **(O)** Relations between the execution index and within-subject differences in the nose poke rate between days 9 & 13 are plotted in the XY-plane. (H–O) $VR+VI_{High}$ N=40; $VR+VI_{Low}$ N=39. **(P, Q)** 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) N=9, (Q) N=8. **(R)** 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

1081 and second devaluation tests are plotted in XY-planes. N=17. **P < 0.01 and ***P < 0.001; N.S., not

1082 significant.

1083



1084 **Figure 2**

1085 *Cortical region-specific plasticity is induced at each training stage*

1086 **(A–F)** Acute brain slices of the anterior cingulate cortex (ACC), lateral orbitofrontal cortex (LOFC)
1087 and medial OFC (MOFC) were prepared within 30 min after the 3rd CRF and 6th VR training. Control
1088 (Ctrl) mice received food restriction but were not trained. AMPA/NMDA ratios were recorded from
1089 pyramidal neurons in the ACC layer 2/3 (A), ACC layer 5 (B), LOFC layer 2/3 (C), LOFC layer 5
1090 (D), MOFC layer 2/3 (E) and MOFC layer 5 (F). (A) N=10–16; (B) N=9–18; (C) N=9–10; (D) N=9–
1091 10; (E) N=9–11; (F) N=9–10 for each group. **(G–L)** Acute brain slices of the ACC, LOFC and medial
1092 OFC were prepared within 30 min after the 4th VI training session. The VR+VR group was introduced
1093 as a control matched for total training days. AMPA/NMDA ratios were recorded from pyramidal
1094 neurons in the ACC layer 2/3 (G), ACC layer 5 (H), LOFC layer 2/3 (I), LOFC layer 5 (J), MOFC
1095 layer 2/3 (K) and MOFC layer 5 (L). The black and blue dashed lines show the mean value of the Ctrl

1096 and VR groups in (A–F) respectively. (G) N=9–12; (H) N=9; (I) N=9–10; (J) N=9–12; (K) N=9; (L)

1097 N=8–9 for each group. **(M–R)** Correlation between the execution index and AMPA/NMDA ratios in

1098 the $VR+VI_{High}$ and $VR+VI_{Low}$ groups. N=18–19. *$P < 0.05$; **$P < 0.01$; N.S., not significant.

1099

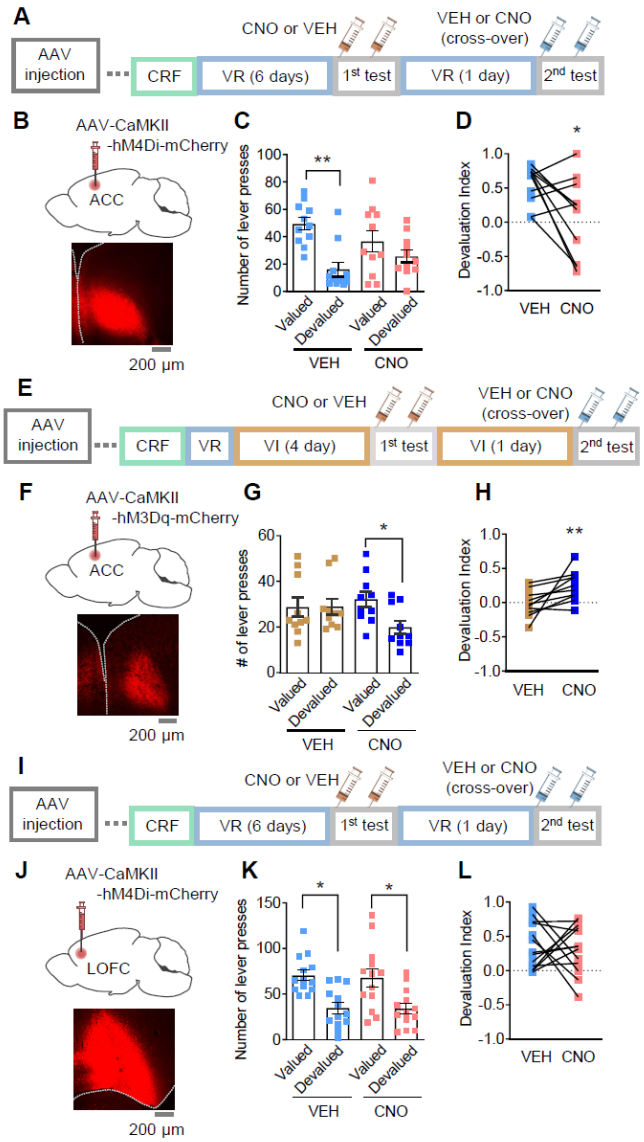


Figure 3

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. N=11. **(D)** Within-subject changes in the devaluation index of (C). **(E)** Mice received injection with AAV-CaMKII-hM3Dq-mCherry in the ACC. The first

1110 devaluation test was performed following the 4th VI training session. After the first test, mice were
1111 re-trained with the VI schedule for one additional day, after which the second devaluation test was
1112 performed. Same as (A), drug injections were conducted in a cross-over design. **(F)** Representative
1113 image of hM3Dq-mCherry expression in the ACC. **(G)** Effects of ACC activation on lever pressing
1114 in the devaluation tests. N=10. **(H)** Within-subject changes in the devaluation index of (G). **(I)** Mice
1115 received injection with AAV-CaMKII-hM4Di-mCherry in the LOFC. The experimental protocol was
1116 same as (A), except for the AAV injection sites. **(J)** Representative image of hM4Di-mCherry
1117 expression in the LOFC. N=13. **(K)** Effects of the LOFC inhibition on lever pressing in the
1118 devaluation tests. **(L)** Within-subject changes in the devaluation index of (K). *P < 0.05; **P < 0.01.
1119

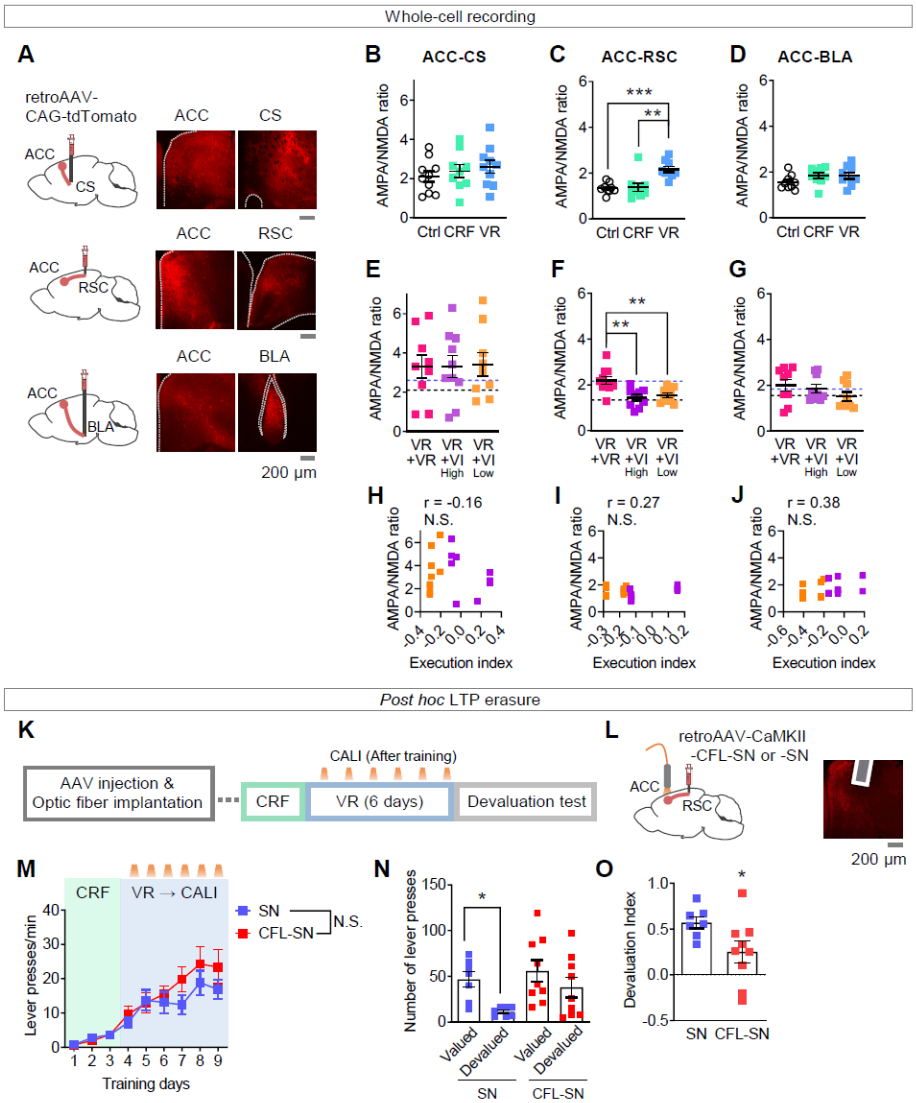


Figure 4

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). (B) N=9–10; (C) N=9; (D) N=9 for each group. **(E–G)** Acute brain slices of the ACC

1130 were prepared within 30 min after the 4th VI training. The VR+VR group was introduced as a control
1131 matched for total training days. AMPA/NMDA ratios were recorded from CS-projecting ACC layer
1132 5 neurons (E), RSC-projecting ACC layer 5 neurons (F) and BLA-projecting ACC layer 5 neurons
1133 (G). Black and blue dashed lines show the mean value of Ctrl and VR groups in (B–D) respectively.
1134 (E) N=9–10; (F) N=9–10; (G) N=9 for each group. **(H–J)** Correlation between execution index and
1135 AMPA/NMDA ratios in the VR+VI_{High} and VR+VI_{Low} groups. N=18–19. **(K)** Schematic diagram of
1136 VR training with CALI. Mice received CALI (illumination with 1–2 mW amber light for 1 min)
1137 within 5-min after each VR training session ended (days 4–9). **(L)** Representative image of CFL-SN
1138 expression and optic fiber implantation in the ACC. retroAAV-CaMKII-CFL-SN or retroAAV-
1139 CaMKII-SN were injected into the RSC and optic fibers implanted into the ACC. **(M)** Lever press
1140 rate during training with CALI. **(N)** Effects of CALI in RSC-projecting ACC neurons on lever
1141 pressing in the devaluation tests. **(O)** Within-subject changes in the devaluation index of (N). SN,
1142 N=7; CFL-SN, N=9. *P < 0.05; **P < 0.01; ***P < 0.001; N.S., not significant.
1143

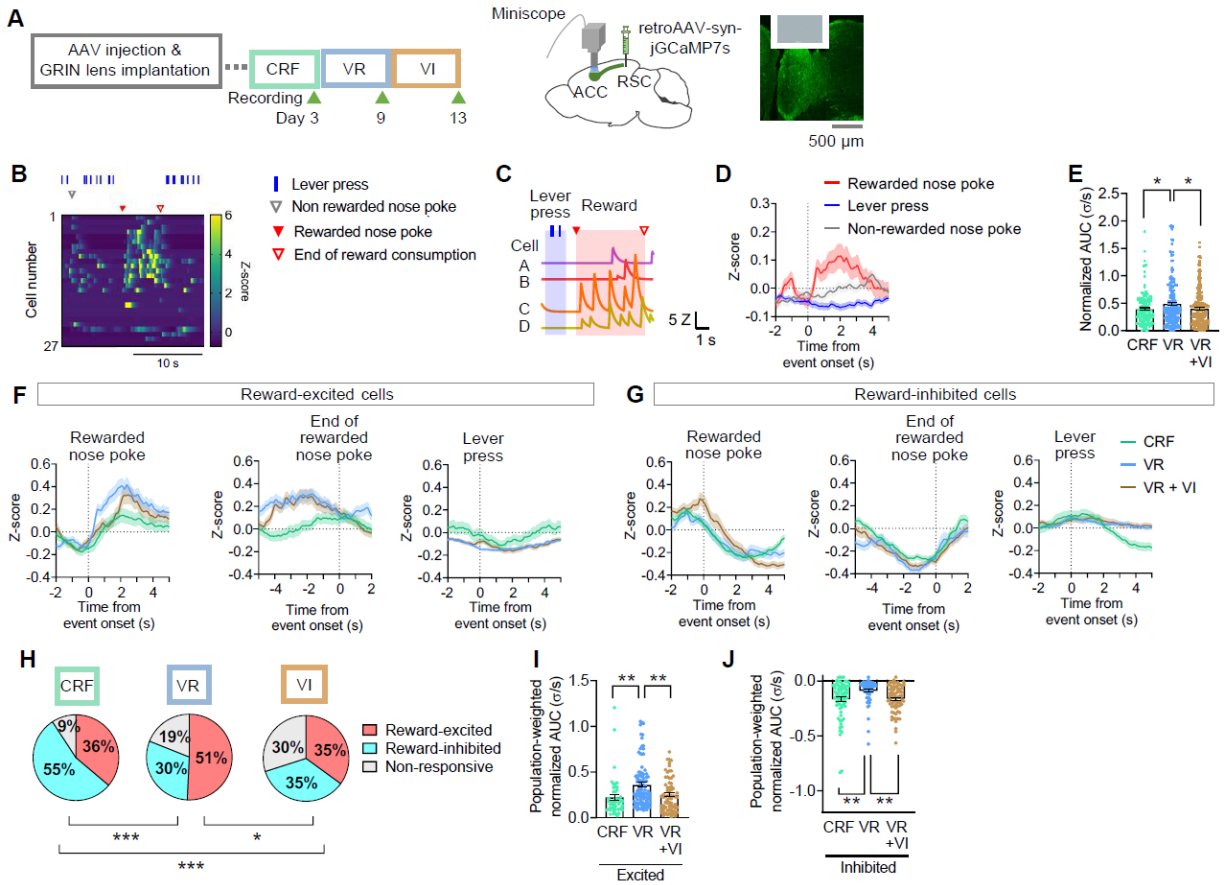


Figure 5

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 (see Methods). **(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. CRF, N=130;

1157 VR, N=195; VR+VI, N=197. **(F,G)** The average temporal dynamics of reward-excited **(F)** and -
1158 inhibited **(G)** cells at the start and end of rewarded nose pokes and lever pressing during each training
1159 stage. **(H)** Fractions of reward-excited and reward-inhibited cells at each training stage. **(I,J)** For
1160 reward-excited and reward-inhibited cells, the population-weighted normalized AUC were calculated
1161 by multiplying normalized AUCs of positive (I) or negative (J) reward-related responses by the
1162 fraction of the reward-excited and reward-inhibited cells. (I) CRF, N=47; VR, N=99; VR+VI, N=69;
1163 (J) CRF, N=71; VR, N=59; VR+VI, N=69. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
1164

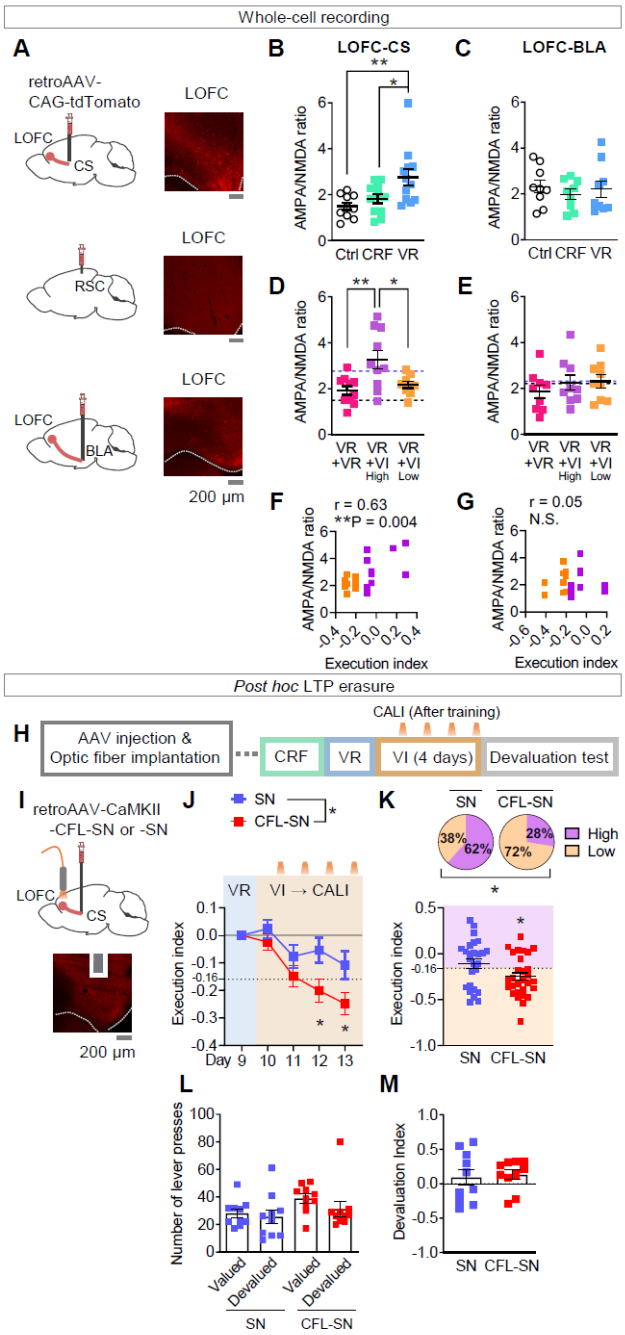


Figure 6

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

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

1172 Acute brain slices of LOFC were prepared within 30 min after the 3rd CRF and 6th VR training.
1173 Control (Ctrl) mice underwent food restriction but were not trained. AMPA/NMDA ratios were
1174 recorded from CS-projecting LOFC layer 5 neurons (B) and BLA-projecting LOFC layer 5 neurons
1175 (C). (B) N=10–12; (C) N=9 for each group. **(D,E)** Acute brain slices of the LOFC were prepared
1176 within 30 min after the 4th VI training. The VR+VR group was introduced as a control matched for
1177 total training days. AMPA/NMDA ratios were recorded from the CS-projecting LOFC layer 5 neurons
1178 **(D)** and BLA-projecting LOFC layer 5 neurons **(E)**. Black and blue dashed lines show the mean value
1179 of the Ctrl and VR groups in (B,C) respectively. (D) N=9–10; (E) N=9 for each group. **(F,G)**
1180 Correlation between the execution index and AMPA/NMDA ratios in the VR+VI_{High} and VR+VI_{Low}
1181 groups. N=18–19. **(H)** Schematic diagram of the two-step training protocol with CALI. Mice received
1182 CALI (illumination with 1–2 mW amber light for 1 min) within 5-min after each VI training (day 10–
1183 13). **(I)** Representative image of CFL-SN expression and optic fiber implantation in the LOFC. The
1184 retroAAV-CaMKII-CFL-SN or retroAAV-CaMKII-SN was injected into the CS and optic fibers were
1185 implanted into the LOFC. **(J)** Execution index during VI training with CALI. **(K)** Execution index
1186 on day 13 and the fraction of VR+VI_{High} and VR+VI_{Low} mice in each group. SN, N=26; CFL-SN,
1187 N=29. **(L)** Effects of CALI in the CS-projecting LOFC neurons on lever pressing in the devaluation
1188 test. SN, N=10; CFL-SN, N=10. **(M)** Within-subject changes in the devaluation index of (L). *P <
1189 0.05; **P < 0.01; N.S., not significant.

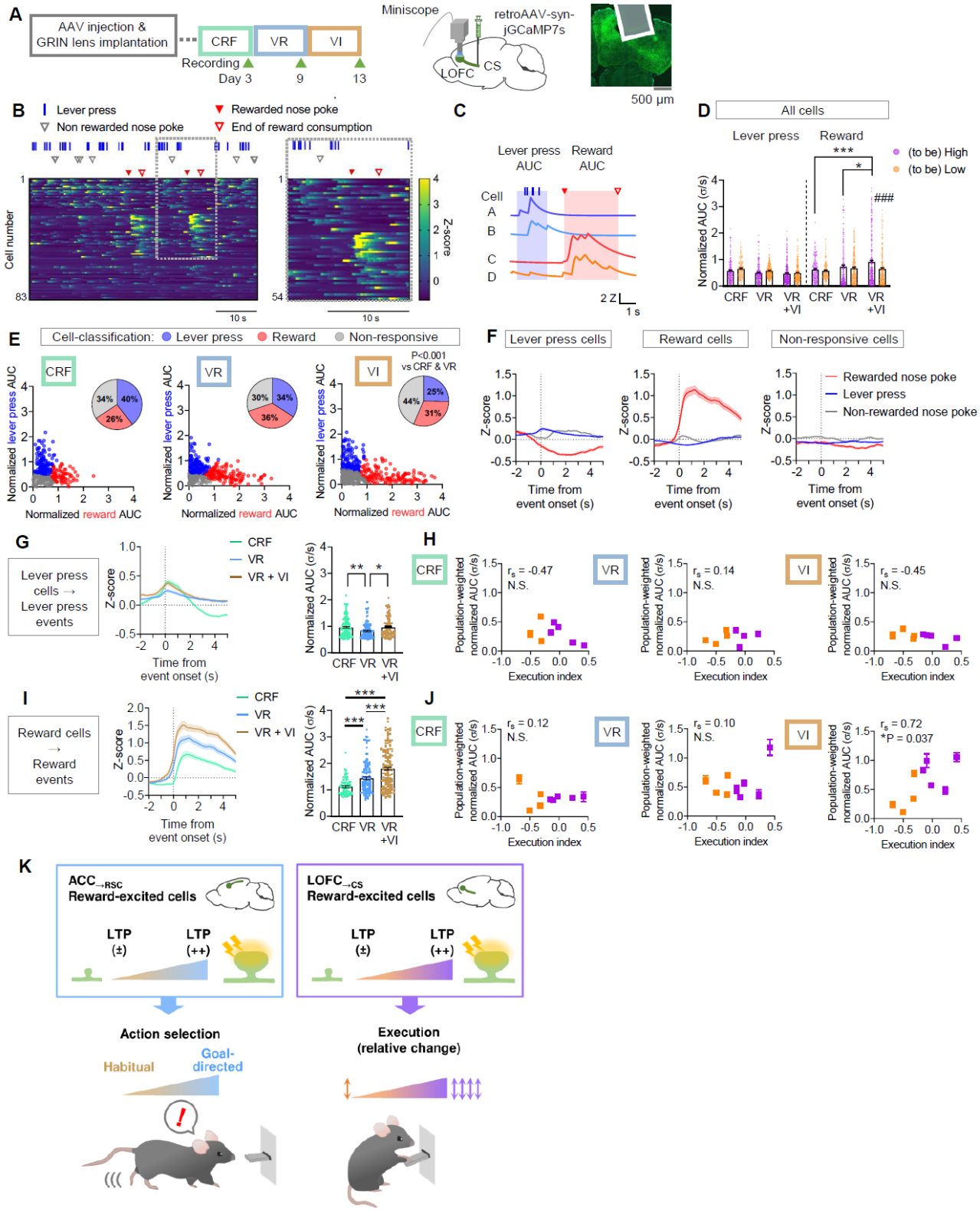


Figure 7

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

1194 and GRIN lens implantation. The retroAAV-syn-jGCaMP7s was injected into the CS and GRIN lens
1195 was implanted into the LOFC. Recordings were performed in the last CRF (day 3), VR (day 9) and
1196 VI (day 13) training sessions. **(B)** Representative heat map of Ca^{2+} transients during VR training (day
1197 9). The right panel shows the enlarged version of the dotted area on the left. **(C)** Representative trace
1198 of Ca^{2+} transients around lever pressing and reward events. Cells A and B were activated during lever
1199 presses, but not at reward consumption. In contrast, cells C and D showed higher activity during
1200 reward consumption than at lever presses. The red and blue highlighted zones show the time windows
1201 for calculation of reward and lever press-related neural responses, respectively (see Methods). **(D)**
1202 The AUC of event-related responses were normalized by the duration of lever press bouts, or reward
1203 consumption. Each dot represents each cell's average of normalized AUCs for lever press or reward
1204 events during training. CRF (to be High), N=166; CRF (to be Low), N=140; VR (to be High), N=179;
1205 VR (to be Low), N=208; $VR+VI_{High}$, N=247; $VR+VI_{Low}$, N=213. **(E)** The XY-planes represent lever
1206 press and reward-related normalized AUCs for each cell. Cells were classified into three subgroups
1207 according to the magnitude and preference of their responses toward lever press and reward events
1208 (see Methods). The pie charts represent fractions of each subgroup. **(F)** Average activity of lever press
1209 cells, reward cells and non-responsive cells during each type of event during VR training. **(G)** Average
1210 temporal dynamics and normalized AUCs of lever press cells to lever press events during each
1211 training stage. CRF, N=121; VR, N=132; VR+VI, N=116. **(H)** Correlation between the execution
1212 index and population-weighted AUCs of the lever press cells to lever press events during each training
1213 stage. CRF, N=121 cells/8 mice; VR, N=132 cells/8 mice; VR+VI, N=116 cells/9 mice. **(I)** Average
1214 temporal dynamics and normalized AUCs of the reward cells to reward events during each training
1215 stage. CRF, N=80; VR, N=138; VR+VI, N=144. **(J)** Correlation between the execution index and
1216 population-weighted AUCs of the reward cells to reward events during each training stage. CRF,
1217 N=80 cells/9 mice; VR, N=138 cells/9 mice; VR+VI, N=144 cells/9 mice. *$P < 0.05$; **$P < 0.01$;
1218 ***$P < 0.001$; 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 system for action selection, and as it weakens, behavior becomes habitual. By contrast, 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.