



Published in final edited form as:

Nat Commun. 2013 ; 4: 2264. doi:10.1038/ncomms3264.

## Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions

Christina M. Gremel<sup>1</sup> and Rui M. Costa<sup>1,2</sup>

<sup>1</sup>Laboratory for Integrative Neuroscience, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, TS-20, 5625 Fishers Ln, Bethesda, USA20892

<sup>2</sup>Champalimaud Neuroscience Programme, Champalimaud Institute for the Unknown, Rua Da Quinta Grande, N6, Oeiras, 2780-156, Portugal

### Abstract

Shifting between goal-directed and habitual actions allows for efficient and flexible decision-making. Here we demonstrate a novel, within-subject instrumental lever-pressing paradigm where mice shift between goal-directed and habitual actions. We identify a role for orbitofrontal cortex (OFC) in actions following outcome-revaluation, and confirm that dorsal medial (DMS) and lateral striatum (DLS) mediate different action strategies. *In-vivo* simultaneous recordings of OFC, DMS, and DLS neuronal ensembles during shifting reveal that the same neurons display different activity depending on whether presses are goal-directed or habitual, with DMS and OFC becoming more—and DLS less-engaged during goal-directed actions. Importantly, the magnitude of neural activity changes in OFC following changes in outcome value positively correlates with the level of goal-directed behavior. Chemogenetic inhibition of OFC disrupts goal-directed actions, while optogenetic activation of OFC specifically increases goal-directed pressing. They also reveal a role for OFC in action revaluation, which has implications for understanding compulsive behavior.

---

We often perform a similar action for different reasons, either to achieve a particular goal at that moment, or because this action has been routinely reinforced and is now habitual<sup>1-4</sup>. Although the development of habits and rules is important for responding rapidly and accurately given a particular stimulus or state, we also encounter circumstances that cause us to re-evaluate the consequences of our actions. An inability to shift between habits and goal-directed actions (“break habits”) may underlie distorted behaviors observed in obsessive compulsive disorder (OCD), addiction and other decision-making disorders<sup>2,3,5-11</sup>.

The neural mechanisms and circuits governing the shift between these two behavioral strategies remain elusive. In the dorsal striatum, which receives vast inputs from most

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Correspondence can be addressed to [gremelc@mail.nih.gov](mailto:gremelc@mail.nih.gov) or [ruicosta@fchampalimaud.org](mailto:ruicosta@fchampalimaud.org).

**Author Contributions.** C.M.G. performed the experiments and analyzed the data. C.M.G. and R.M.C. designed the experiments, designed data analyses, and wrote the paper.

**Competing Financial Interests.** The authors have no competing financial interests to declare.

cortices<sup>12–14</sup>, the dorsal medial striatum (DMS) is necessary for goal-directed actions; lesions or inactivation of DMS render actions habitual instead of goal-directed<sup>15</sup>. Conversely, the dorsal lateral striatum (DLS), is necessary for habitual actions; lesions or temporary inactivation of DLS bias behavior towards goal-directed actions<sup>16,17</sup>. Furthermore, the balance between habits and goal-directed behavior is impaired in diseases such as obsessive-compulsive disorder<sup>8</sup>, in which the orbital frontal cortex (OFC) is dysfunctional<sup>18–20</sup>. This suggests that shifting between goal-directed and habitual actions could involve dynamic interactions between the corticostriatal circuits that underlie these individual behavioral strategies. However, how behavioral shifting is implemented is unknown.

One possibility would be that these action strategies are encoded by different neuron ensembles in corticostriatal circuits, and a shift in behavior would correspond to a shift of activity between neurons controlling goal-directed actions and neurons controlling habits. Another possibility would be that action strategies are concurrently encoded in the same neuronal ensembles in these circuits, and a shift between goal-directed actions and habits would correspond to a shift of activity in the same neurons as the different circuits compete to gain control over behavior output.

To disambiguate between these possibilities, we demonstrate a novel instrumental task where the same mouse would readily shift between performing a similar action for the same reward using either a goal-directed or a habitual strategy. Our results from experiments using functional lesions, *in vivo* recordings during action learning and reevaluation, chemogenetic as well as optogenetic stimulation, suggest that shifts in activity of the same corticostriatal neuronal ensembles correspond to and can cause shifts between goal-directed and habitual actions.

## Results

### Mice readily shift between goal-directed actions and habits

Paradigms to examine isolated goal-directed and habitual actions have been developed in humans and rodents, and outcome reevaluation procedures examining control by the current expected value are commonly used to operationally distinguish these two behavioral strategies<sup>10,11</sup>. We designed a novel self-paced instrumental task where individual mice readily shifted between performing goal-directed actions and habits. We took advantage of different contextual cues to differentiate between commonly used random ratio (RR) and random interval (RI) reinforcement schedules that bias towards the generation of goal-directed versus habitual actions, respectively<sup>1,2,4,2110</sup> (Methods). We trained mice to press the same manipulandum (a lever placed in the same location) for the same reinforcer, using both RI and RR schedules of reinforcement (Fig.1A, Methods). Mice were initially trained to lever press on a continuous reinforcement (CRF) schedule, with the potential to earn 5, 15, and 30 rewards across 3 days. Then, mice underwent two days of RI30 (reinforcement follows the first press after 30 seconds on average has passed) and RR10 (reinforcement follows on average the 10<sup>th</sup> lever press) training, followed by four days of RI60 and RR20 training.

Mice ( $n = 10$ ) similarly increased pressing rate across days of training in both schedules (main effect Training day:  $F(8, 144) = 20.15, p < 0.001$ ), with mice making slightly more lever-presses during RR training (interaction and main effects:  $F_s' > 3.2, p_s' < 0.01$ ) (Day 3 and 5 Bonferroni-corrected  $p_s' < 0.05$ , Fig.1b, Supplementary Fig.S1a). Importantly, mice earned similar numbers of rewards, earned rewards at a similar rate, and made a similar number of head entries into the food port between RI and RR schedule training (no interaction or Schedule main effect  $F_s' < 0.9, p_s' > 0.05$ ; main effect Training Day:  $F_s' > 2.40, p_s < 0.05$ ) (Supplementary Fig. S1b-d). We also verified that in RI schedule training there was no scalloping in responding<sup>22</sup> (Supplementary Fig.S1g, h). Further, we verified that the distribution of inter-reward intervals was the same between RI and RR schedules (Supplementary Fig. S1i, j), together suggesting that RI and RR schedules produced very similar patterns of lever-press behavior.

Since the action strategy employed (goal-directed or habitual) cannot be elucidated during training<sup>23</sup>, we probed the degree to which an action in each training context was goal-directed or habitual during a brief (5 min) outcome revaluation test. We measured the number of non-reinforced lever-presses in each context following sensory-specific satiation with either the outcome earned by lever pressing (devalued state), or a control outcome given daily in the home-cage (valued state) (Methods). We observed that mice reduced lever-pressing only in the RR context, but not RI context (Repeated measures ANOVA (Revaluation state  $\times$  Schedule) interaction:  $F(1, 18) = 4.51, p < 0.05$ ) (RR context: Bonferroni-corrected  $p < 0.001$ ) (Fig.1c) (Supplementary Fig. S1e). Further one-sample t-tests of normalized lever-pressing against chance (0.5) showed that only in the RR context did lever-pressing significantly differ, with more pressing in the Valued state, and less pressing in the Devalued state (RR context:  $t_s' > 4.29, p_s' < 0.002$ ; RI context:  $t_s' < 1.27, p_s' > 0.2$ ). These data show that lever-pressing in the same mouse was sensitive to outcome revaluation in the RR but not the RI schedule training context, and indicate that contextual information can induce mice to readily shift between executing a similar action in a goal-directed versus habitual manner. Non-rewarded head entries to the food port reduced following outcome revaluation in both previously RI and RR trained contexts ( $F(1, 18) = 6.11, p < 0.05$ ) (RI context:  $t_s'_{10} = 2.33, p_s' < 0.05$ ) (Fig. 2d).

### Corticostriatal circuits controlling action strategies

We next examined the contribution of DMS and DLS to the shift between goal-directed and habitual actions. Excitotoxic lesions to either the DMS or DLS in mice (final  $n = 5-9$  per group) (Supplementary Fig. S2a, Methods) did not grossly impair acquisition of lever-press behavior under RI and RR schedule training (no interaction or Schedule main effect, main effect Training day:  $F(16, 128) = 28.75, p < 0.0001$ ) (Fig.1e, f) (Supplementary Fig.S3b, c, d). During outcome revaluation testing, sham mice reduced responding in the RR but not RI contexts following outcome revaluation (Schedule  $\times$  Revaluation state) interaction:  $F(1, 12) = 2.94, p = 0.07$ ; RR context Bonferroni-corrected  $p < 0.01$ ) (no main effects) (one-sample t-test (0.5) Valued and Devalued states: RI context:  $t_s' < 1.27, p_s' = 0.06$ ; RR context:  $t_s' < 4.45, p_s' < 0.002$ ) (Fig.1g). However, during testing we found that DMS-lesioned mice were always habitual and insensitive to outcome revaluation in both training contexts (Schedule  $\times$  Revaluation state, no interaction or main effects:  $F_s' < 0.95, p_s' > 0.1$ ) (one-sample t-test

(0.5) on Valued and Devalued states: RI and RR contexts  $ts'_4 < 0.80$ ,  $ps' > 0.4$ ). Conversely, mice with DLS lesions reduced lever-pressing following outcome revaluation, and were goal-directed in both training contexts (no interaction or main effect Schedule, main effect Revaluation state:  $F_{1, 10} = 11.29$ ,  $p < 0.01$ ) (RI and RR Schedule Bonferroni-corrected  $ps' < 0.01$ ) (one-sample t-test (0.5) on Valued and Devalued states: RI and RR contexts  $ts'_5 > 2.53$ ,  $ps' < 0.05$ ) (Fig. 1g, Supplementary Fig. S3g). These results show that within-subject shifts are also controlled by dorsal striatal subregions<sup>9</sup>, and demonstrate that impediment to use the circuit involved in a particular action strategy results in a bias towards the use of the remaining intact circuit for action execution, suggestive of parallel encoding of both action strategies.

Since OFC has been implicated in various cue-related behaviors modulated by changes in expected value<sup>25–39</sup>, and OFC dysfunction has been linked to obsessive-compulsive disorder<sup>18–20</sup>, we examined its role in shifting between goal-directed and habitual actions. The OFC modulates medial striatum through direct projections<sup>12,40,41</sup> (Supplementary Fig. S12b), and indirectly through connections with striatal projecting cortical areas, basolateral amygdala and ventral tegmental/substantia nigra (*pars compacta*)<sup>12</sup> (Supplementary Fig. S12c, d), nuclei known to contribute to instrumental actions<sup>42–44</sup>. We examined behaviors only in mice with localized more lateral versus medial OFC lesions<sup>45</sup> not affecting neighboring cortices (final sham  $n = 7$ , OFC  $n = 5$  per group, excluded  $n = 5$  for extension of lesion to neighboring regions) (Supplementary Fig. S2b). OFC lesions did not affect acquisition of lever-press behavior in the RI schedule (Training day  $\times$  Lesion group, no interaction or main effect Lesion group, main effect Training day:  $F_{8, 56} = 10.69$ ,  $p < 0.001$ ) (Fig. 1h, i, Supplementary Fig. S4b, although there were fewer lever presses on the last two days of RR schedule training). Although visual inspection of the data suggested mice with OFC lesions had higher response rates under RI than RR schedules, this was non-significant ( $F < 1.04$ ,  $p > 0.4$ ). Further, no effects of OFC lesion were observed on the number of lever-presses made, rewards earned, rate of rewards earned, or head entry behavior in either schedules (no interaction or Lesion group main effect) (main effect Training day:  $F_s' > 1.96$ ,  $ps' > 0.06$ ) (Supplementary Fig. S4b–e).

OFC-lesioned mice did not reduce lever-pressing in either context following outcome revaluation (no interaction Schedule  $\times$  Devaluation state, or main effects:  $F_s' < 0.50$ ,  $ps' > 0.05$ ) (one-sample t-test (0.5) on Valued and Devalued states: RI and RR contexts:  $ts'_4 < 1.09$ ,  $ps' > 0.3$ ), while Sham mice shifted between habitual and goal-directed actions (Schedule  $\times$  Devaluation state interaction:  $F_{1, 8} = 8.53$ ,  $p < 0.05$ ) (Sham RR context: Bonferroni-corrected  $p < 0.01$ ) (one-sample t-test on Valued and Devalued states: RI context  $ts'_6 = 1.09$ ,  $ps' < 0.35$ ; RR context  $ts'_6 > 3.90$ ,  $ps' < 0.05$ ) (Fig. 1j, Supplementary Fig. S4f). Similar consumption between groups suggested no difference in outcome valuation (Supplementary Fig. S4h). Further, the impairment observed in OFC-lesioned mice was not caused by an inability to discriminate between contexts. Separate groups of OFC-lesioned mice trained independently on either RI ( $n = 10$ ) or RR ( $n = 11$ ) schedules of reinforcement (Methods)<sup>46</sup> still showed intact habitual actions (Supplementary Fig. S4i–o), but disrupted goal-directed actions (Supplementary Fig. S4p–v). There was no correlation between the response rate or reinforcement rate, on the last day of training and the revaluation indices

(methods) for each mouse in either OFC-lesioned ( $r_{33} < 0.13$ ) or Sham mice ( $r_{26} < 0.16$ ). Together this suggests that OFC is critical for the sensitivity of instrumental actions to changes in outcome value.

### Concurrent encoding of goal-directed and habitual actions

Using multi-site multi-electrode recordings *in vivo* (Methods, Supplementary Fig. S5g-k) we recorded the activity of the same DMS, DLS, and OFC neurons in each mouse during both RI and RR schedule training ( $n = 8$  mice; Fig. 2a, b) (Supplementary Fig. S5). Recorded neurons showed similar baseline firing rates between training contexts (Supplementary Fig. S6 a-b, e-f, i-j). As in other studies, we found evidence of changes in firing rate of DMS, DLS, and OFC neurons around the lever press<sup>49,50</sup> ( $\pm 2$  sec) during both RI and RR schedule training (Fig. 2c, g, and k), with phasic increases in activity typically preceding lever pressing (Supplementary Fig. S5i-k).

Previous findings using a cued-task have suggested similar engagement of DMS and DLS circuits<sup>51,52</sup>. Using training schedules to directly bias the generation of instrumental habitual or goal-directed actions, we observed similar proportions of lever-press related neurons between RI and RR schedules in DMS and DLS, as well as OFC circuits (per mouse, Fig 2d, h, and l) ( $p > 0.05$ ). Further, we observed fairly similar proportions of up and down-modulated neurons that increased or decreased their firing rate, respectively, during lever-press behavior, (Supplementary Fig. S7).

In the within-subject design, we can examine activity changes in the same neuron during lever-press behavior under schedules biasing goal-directed and habitual actions. Changes in lever-press related activity could represent the same neuron modulating activity under both RI and RR schedules (Both-schedule neurons), or Schedule-specific neurons that modulate their activity specifically during pressing in either the RI or the RR training context. We found a larger proportion of Both-schedule neurons than Schedule-specific neurons in DMS, DLS, and OFC during RI and RR schedule training (DMS  $\chi^2 = 22.60$ ,  $p < 0.0001$ ; DLS  $\chi^2 = 7.12$ ,  $p = 0.07$ ; OFC  $\chi^2 = 13.49$ ,  $p < 0.004$ ) (Fig. 2e, i, and m). Given this finding, it could be that the same neurons (Both-schedule neurons) show different rate modulation during lever-pressing depending on the training schedule. We used a modulation index to examine the degree to which each Both-schedule neuron was differentially modulated during lever-pressing under RR and RI schedules (Supplementary Fig. S6c, g, and i), [(RR modulation rate – RI modulation rate)/(RR modulation rate + RI modulation rate)].

We found evidence in all three areas that some Both-schedule neurons showed stronger modulation in one context or the other (RI vs. RR). Averaging across Both-schedule neurons in DMS and DLS did not reveal modulation differences between RR and RI schedules; however, there was a training-induced shift in OFC modulation from Day 1 to 6 ( $t_{36} = 3.66$ ,  $p < 0.001$ ), with initially greater modulation in RR on Day 1 ( $t_{19} = 2.54$ ,  $p < 0.05$ ), to greater modulation in RI on Day 6 ( $t_{17} = 3.3$ ,  $p < 0.01$ ). Careful inspection of the index for DLS Both-schedule neurons revealed two distinct populations on Day 6, and analyses showed a non-Gaussian distribution on Day 6 ( $K2 = 6.04$ ,  $p < 0.05$ ) (Fig 2j). DLS Both-schedule neurons that increased firing rate during lever-pressing showed a negative modulation index score ( $-4.41 \pm 0.76$ ,  $t_5 = 5.82$ ,  $p < 0.01$ ). DLS Both-schedule neurons that decreased firing

rate during lever-press behavior showed a positive index score ( $4.65 \pm 0.34$ ,  $t_6 = 13.70$ ,  $p < 0.001$ ). This suggests that DLS neurons become more inhibited with continued goal-directed training, and more active during continued habit training.

### Shifts in neural modulation correspond to shifts in behavior

The findings presented above support the hypothesis that acquisition of goal-directed and habitual actions occurs in parallel in these circuits; and that often the same neurons are involved in both types of action, albeit differently modulated. This raises the possibility that the shift between goal-directed and habitual actions is reflected in differences in the modulation of Both-schedule neurons. To test this hypothesis, we examined the lever-press related change in firing rates in DMS, DLS, and OFC neurons during outcome revaluation testing (Fig 3a, b and 4a, b) (Supplementary Fig. S8) ( $n = 6$  mice) (Schedule $\times$ Revaluation state interaction:  $F_{1, 28} = 6.36$ ,  $p < 0.05$ ) (RR context: Bonferroni-corrected  $p < 0.001$ ) (RI context:  $ps' > 0.05$ ) (one-sample test (0.5) for Valued and Devalued states: RI context:  $ts'_7 < 0.24$ ,  $ps' > 0.8$ ; RR context:  $ts'_7 > 3.07$ ,  $ps' < 0.05$ ). While normally goal-directed and habitual processes most likely contribute jointly to action control, outcome revaluation procedures promote goal-directed actions and habits to compete for action control<sup>10</sup>.

We first investigated the absolute change in rate modulation in DMS, DLS, and OFC neuron ensembles during lever-press behavior following outcome revaluation (Fig. 3c, f, and i). Overall, there was a trend in OFC and DMS towards greater rate modulation in the previously trained RR vs. RI contexts (Fig. 3d, and j) (OFC  $t_{65} = 2.77$ ,  $p = 0.07$ ; DMS  $t_{48} = 1.78$ ,  $p = 0.09$ ), but not in DLS ( $t_{30} = 0.14$ ,  $p = 0.10$ ) (Fig 3g). To examine the contribution of changes in the firing rate of the same neuron to differences observed in modulation rate above, we next examined the modulation rate of Both-schedule neurons in these circuits (Fig. 3d, j, g). There was greater rate modulation in the previous RR trained context of OFC Both-schedule neurons ( $t_{17} = 2.28$ ,  $p < 0.05$ ), and DMS Both-schedule neurons (although to a lesser extent,  $t_{16} = 2.0$ ,  $p = 0.06$ ), (Fig. 3d, and j). This was not observed for DLS Both-schedule neurons or for Schedule-specific neurons ( $ps' > 0.05$ ) (Supplementary Fig. S9).

Next, we examined whether these changes in rate modulation of Both-schedule neurons really reflect the shift in behavior following outcome revaluation. We correlated the modulation index in the devalued state, with a revaluation index assessing the sensitivity of lever-press behavior to changes in value in previously trained RI and RR contexts.

We found that, in the devalued state, the relative modulation of Both-schedule neurons in DMS and OFC in the previously RR versus RI trained contexts, positively correlated with the degree of goal-directed behavior (Fig. 3c and i). That is, in the devalued state, the stronger the modulation of the same DMS and OFC neuron was during pressing in RR vs. RI, the more sensitive goal-directed lever pressing was to changes in outcome value. Interestingly, the converse tendency was observed in DLS. However, no significant correlations were observed for habitual actions in the RI context in DMS, DLS, or OFC. Additionally, we did not observe a similar relationship between DMS, DLS, and OFC neurons specific to the RI or RR schedule and behavior (Supplementary Fig. S9).



In contrast, we did not observe differences in rate modulation of DMS and OFC neurons between RI and RR contexts in the valued state where action value remains high ( $ps' > 0.1$ ) (Fig 4b, e, and h). Also, DMS and OFC Both-schedule neurons ensembles showed similar rate modulation between RI and RR contexts ( $ps' > 0.1$ ) (Fig. 4C, and i), with no correlation between modulation index and sensitivity to outcome revaluation (Fig. 4d, j). However, when action value was high, DLS neuron ensembles showed less rate modulation in the RR than RI context (Fig. 4e) ( $t_{49} = 1.98, p = 0.05$ ). Further, the less DLS Both-schedule neurons modulated firing rate in RR versus RI, the more sensitive lever-pressing in the RR context was to outcome revaluation (Fig. 4g). Together, these findings suggest that the sensitivity of actions to changes in outcome value during goal-directed behavior is related to stronger modulation of OFC and DMS neurons, and weaker modulation of DLS neurons, in the RR versus the RI contexts.

### OFC conveys information about changes in action value

These findings raise the hypothesis that reductions in goal-directed actions from the Valued to the Devalued state are related to changes in the overall modulation of OFC, DMS, and DLS for each subject. To examine this, we first calculated the change in neural ensemble modulation (Both-schedule and Specific-schedule neurons) between Valued and Devalued states in OFC, DMS and DLS for each mouse, and for both the RI and RR contexts (Supplementary Fig. S10). Next, we correlated this change in modulation between Valued and Devalued states with the sensitivity to outcome revaluation. A striking positive correlation was observed in OFC ( $p = 0.01$ ) (less in DMS,  $p = 0.08$ ), revealing that larger differences in OFC neural ensemble modulation between value states corresponded to greater sensitivity of actions to changes in outcome value (in the RR context, Figure 5a). i.e. for each mouse, less OFC modulation in the Devalued vs. the Valued state correlated with a stronger reduction in pressing following devaluation. This was not observed for habitual actions (RI context) for any area (Figure 5a, b).

These results provide additional evidence suggesting that OFC ensembles are conveying information about action-value. To test this hypothesis, we changed the activity of OFC projection neurons during outcome revaluation testing. We first reduced the activity of OFC projections using a chemogenetic approach with designer receptor (DREADD) exclusively activated by the designer drug clozapine *N*-oxide (CNO)<sup>53,54</sup> (Methods). A cre-dependent viral vector expressing G<sub>i</sub>-coupled hM4D<sub>i</sub> DREADD receptors was bilaterally coinjected into OFC with either a virus expressing Cre recombinase under the control of the CaMKII $\alpha$  promoter (restricting Cre expression to pyramidal cells) ( $n = 10$ ; excluded  $n = 2$ ), or a control GFP virus (no DREADD expression) ( $n = 11$ ) (Supplementary Fig. S11). Mice trained concurrently on RI and RR schedules of reinforcement were given systemic injections of the synthetic agonist for hM4D<sub>i</sub> CNO (1 mg/kg) 1 hr prior to outcome revaluation testing leading to a reduction in OFC activity (Supplementary Fig. S11). Inhibition of OFC projection neurons via CNO activation of hM4D<sub>i</sub> receptors disrupted outcome revaluation in the RR context, with mice pressing similarly between valued and devalued states (no interaction or main effects:  $F_s' < 1.79, ps' > 0.1$ ) (one-sample t-test (0.5) for Valued and Devalued states: RI and RR contexts:  $ts'_9 < 0.34, ps' > 0.05$ ) (Figure 5c). As shown before, in control mice, devaluation resulted in a significant reduction in lever-press

behavior specifically in the RR context (Schedule×Revaluation state interaction:  $F_{1, 20} = 3.17, p = 0.07$ ; main effect Revaluation state:  $F_{1, 20} = 14.34, p < 0.01$ ) (RR context: Bonferroni-corrected  $p < 0.01$ ) (RI context:  $p > 0.1$ ) (one-sample t-test (0.5) for Valued and Devalued states: RI context:  $t_{s'_{10}} < 1.27, p_{s'} > 0.2$ ; RR context:  $t_{s'_{10}} > 4.55, p_{s'} < 0.01$ ), indicating that CNO administration had no effect on the shift between goal-directed and habitual actions in control mice. These findings suggest that reducing OFC projection neuron activity during outcome revaluation testing prevents changes in expected outcome value from influencing action performance.

Next, we used an optogenetic approach to selectively activate OFC projection neurons during outcome revaluation testing (Supplementary Fig. S12). Since lesion, DREADD and *in vivo* recording data suggest that OFC activity is not involved in habitual actions, optogenetic activation of OFC projections should not impact habitual actions. In contrast, lesions and DREAD-induced inactivation of OFC disrupted goal-directed actions and there was less OFC lever-press related activity in the Devalued compared to the Valued state, suggesting that the reduced pressing observed following outcome devaluation in the RR context is related to this shift in OFC activity. This leads us to predict that optogenetic stimulation of OFC would increase pressing in the devalued state for goal-directed actions (where action value low), but not in the valued state (where action value is high).

Following injection of a virus expressing channelrhodopsin-2 under the control of the CaMKII $\alpha$  promoter (restricting expression to pyramidal cells)<sup>55</sup> into OFC ( $n = 6$ ) (Supplementary Fig. S12a, e), we concurrently trained mice on RI and RR schedules. We then optically stimulated OFC neurons in both contexts during outcome revaluation testing (i.e. during both revaluation states) (5 ms pulses at 10 Hz) during 5 minutes (light-on), and compared the behavior of the animals to 5 minutes of Light-off in the same sessions. We found that *in-vivo* bilateral stimulation of OFC projection neurons following decrease in outcome value was sufficient to increase lever pressing during this state (Devalued state (Light×Schedule):  $F_{1, 12} = 14.87, p < 0.01$ ) (Fig. 6i). Optogenetic stimulation of OFC did not increase pressing in the Valued state (Valued state  $F < 0.03, p > 0.05$ ) (Fig. 5d) (Supplementary Fig.S12f-j), showing that this manipulation does not just increase the action of pressing, and suggesting that it does increase action value after devaluation. Furthermore, optogenetic stimulation did not alter habitual actions: photostimulation of OFC increased the frequency of lever pressing specifically in the RR context (biasing devalued conditions towards valuation, Bonferroni-corrected  $p < 0.01$ ), but not in the RI context ( $p > 0.05$ ) (Supplementary Fig.S12i-k). These results confirm that changes in OFC activity are related to changes in the performance of goal-directed actions, and provide further evidence that OFC can convey information about action value.

## Discussion

By investigating the activity of the same neurons in corticostriatal circuits as mice performed both goal-directed and habitual actions, we provide evidence that competing orbitofrontal and striatal circuits control context-induced shifts between habitual and goal-directed actions.



We observed that the shifts in activity of the same orbitofrontal and dorsomedial striatal ensembles during outcome revaluation correlated with the degree of goal-directedness, and strikingly, not with execution of habits. These results suggest that although during habitual actions neurons did change activity in relation to outcome-revaluation, the behavior of the animals was independent of the strength of this change. They also suggest that shifting back to goal-directed actions after habits are established corresponds to a dynamic shift in the activity of corticostriatal ensembles, as revealed by greater modulation of DMS and OFC, along with less modulation of DLS, during the performance of goal-directed pressing versus habitual responding.

Finally, we observed that the more lateral OFC is necessary for a shift to goal-directed actions following outcome revaluation. Our findings using lesions, recordings during outcome revaluation, and chemogenetic and optogenetic manipulations directly demonstrate a role for OFC in the balance between goal-directed actions and habits and suggest OFC may be conveying information related to action value. This contrasts with previous findings suggesting a stronger role for OFC in stimulus outcome relations than action outcome relations<sup>47</sup>. One possibility is that the single action to single outcome design used here is more receptive to changes in action-outcome contingency<sup>57</sup>, hence allowing for a shift to habitual actions following disruptions to cortical circuits underlying goal-directed actions. It could also be that inhibiting a single action following devaluation recruits different neural mechanisms than the choice behavior between two outcomes (albeit one devalued) observed following training with two actions and two outcomes.

These results have important implications for understanding neuropsychiatric disorders where the balance between habits and goal-directed actions is disrupted, such as obsessive compulsive disorder<sup>8</sup>. It will be important to determine whether OFC use of outcome value to guide actions is through direct OFC projections to dorsal striatum<sup>12,40,41</sup>, or through indirect projections, (for example through OFC modulation of dopaminergic firing<sup>37</sup> during outcome revaluation). These findings are also important for understanding the execution of and the transition between goal-directed and habitual actions necessary for daily life, which are seemingly impaired in addiction and other decision-making disorders.

## Methods

### Animals

All experiments involved male C57Bl/6J mice at least 7 weeks of age (The Jackson Laboratory, Bar Harbor, ME), and were approved by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the NIAAA Animal Care and Use Committee and done in accordance with NIH guidelines.

### Lesions

0.2  $\mu$ l of N-methyl-D-aspartic acid (NMDA) was infused at a rate of 60 nl/min (via Hamilton syringe) to induce excitotoxic lesions of the DMS (B: AP 0.5mm, L  $\pm$  1.5mm, and V  $-$ 2.5mm from skull) or DLS (B: AP 0.5 mm, L  $\pm$  2.65 mm, and V  $-$ 3.0 mm from skull). Ibotenic acid 0.3  $\mu$ l (10 mg/ml) was infused (via pump, Razel, Scientific) (0.1  $\mu$ l/min) to

induce excitotoxic lesions of the OFC (B: AP 2.7 mm, L  $\pm$  1.75 mm, and V  $-2.25$  mm from dura). For Sham mice, injectors were lowered to the target site but no infusion was given. Mice were allowed to recover for at least 10 days before the start of behavioral procedures. Mice were perfused and brains post-fixed with 4% w/v paraformaldehyde, with lesion placement identified through Nissl staining of 50  $\mu$ m brain slices. Only mice with lesions located with DMS, DLS, or OFC (see Supplementary Figure S2) were included. (Final *n*'s: Striatal Sham lesion = 7, DMS lesion = 5, DLS lesion = 6; OFC Sham = 8–10, OFC lesion = 5–11).

### Chemogenetic inhibition of OFC

For chemogenetic inhibition of OFC projection neurons, cre-inducible AAV-hSyn-DIO-hM4Di-mCherry (Gene Therapy Vector Core at University of North Carolina) was infused bilaterally into OFC (same coordinates as above) with either AAV2/9.CamKII.HI.GFP-Cre or AAV2/ or AAV2/9. GFP virus (University of Pennsylvania vector core) (100 nl/side for each virus). Three weeks following injection, hM4Di mice (*n* = 10) and control mice (*n* = 11) were trained on the within-subject design. During outcome reevaluation testing, mice were given a 1 h pretreatment with clozapine-*N*<sub>1</sub>-oxide (CNO) (1 mg/kg)(10 ml/kg) before operant procedures. To confirm hM4Di activity, we implanted an electrode array at the site of virus infusion. Firing rate of OFC neurons was assessed 1 hr after CNO injection relative to the preceding drug-free baseline-firing rate (Supplementary Fig. S11). Virus spread was assessed under a fluorescence microscope

### Optogenetic activation of OFC

For optogenetic activation of OFC projection neurons, AAV2/9. CamKIIChR2-YFP<sup>S5</sup>(Stanford-Deisseroth lab) (200–300 nl/side) was infused bilaterally into OFC (same coordinates as above) and bilateral optic fiber ferrules were implanted (V  $-2.35$  mm from dura) in OFC. Five weeks following injections, mice (*n* = 6) were trained on the within-subject design. During outcome reevaluation testing, after pre-feeding mice were lightly anesthetized (isofluroane) and connected with a ceramic sleeves to a 473-nm laser via fiber optic rotary joint to optical fibers (200  $\mu$ m core diameter) that was controlled by a Master8 stimulator to deliver 5 ms pulses at 10 Hz (<5mW power at the tip of the fiber). To confirm optogenetic activation of OFC neurons, in a subset of mice (*n* = 2), we attached a fiber optic ferrule to the side of an electrode array to record neural activity at the site of stimulation. We assessed light-activation of OFC neurons in both anesthetized and awake-behaving preparations (Supplementary Fig. S12). AAV2/9. CamKIIChR2-YFP spread and ferrule placement was assessed under a fluorescence microscope.

### Behavioral Procedures

Mice were placed in operant chambers in sound attenuating boxes (Med-Associates, St. Albans, VT) in which they pressed a single lever (left or right) for an outcome of either regular “chow” pellets (20 mg pellet per reinforcer, Bio-Serve formula F05684) or sucrose solution (20–30  $\mu$ l of 20% solution per reinforcer). The other outcome was provided later in their home-cage and used as a control for general satiation in the reevaluation test. Before

training commenced, mice were food restricted to 90% of their baseline weight at which they were maintained for the duration of experimental procedures.

For the within-subject design, training was conducted as follows: each day mice were trained in two separate operant chambers distinguished by contextual cues (black and white striped walls vs. clear plexiglass). For each mouse, the order of schedule exposure, lever position and the outcome obtained upon lever press were kept constant across contexts. However, mice were counterbalanced for context, schedule order, lever position, and outcome earned. Each training session commenced with illumination of the house light and lever extension, and ended following schedule completion or after 90 min with the lever retracting and the house-light turning off.

On the first day, mice were trained to approach the food magazine (no lever present) in each context on a random time (RT) schedule, with a reinforcer delivered on average every 60 sec for a total of 15 min. Next, mice were trained in each context on continuous reinforcement schedules (CRF), where every lever-press made was reinforced, with the possible number of earned reinforcers increasing across training days (CRF5, 15, 30) (recording mice took on average  $6 \pm 1$  days of CRF training (CRF5, 15, 30x4). After acquiring lever-press behavior, mice were trained on random interval (RI) (RI30 2 days/RI60 4 days) and random ratio (RR) (RR10 2 days/RR20 4 days) schedules of reinforcement, with schedules differentiated by context, with the possibility of earning 15 reinforcers in each context or until 90 min had elapsed.

Outcome revaluation testing occurred across two consecutive days as previously described (28). In brief, on the **valued day**, mice had *ad libitum* access to the home-cage outcome for 1 h before serial brief non-reinforced test sessions in the previous RI and RR training contexts. On the **devalued day**, mice were given 1 h *ad libitum* access to the outcome previously earned by lever-press, and then underwent serial non-reinforced test sessions in each training context. Order of context exposure during testing was the same as training exposure, with order of revaluation day counterbalanced across mice. Tests in each context were either 10 min (recording and virus mice) or 5 min (all lesion mice) in duration.

For mice in the between schedule (RI or RR training) lesion experiment, training and devaluation testing proceeded exactly as for mice in the lesion experiment using the within subject design (RI and RR), except that mice were only trained on the RR or RI schedule in one context<sup>46</sup>. Additionally, to equate the total number of possible reinforcers earned between lesion experiments, mice had the opportunity to earn 30 reinforcers or until 90 min had elapsed during RI or RR training.

### ***In vivo* extracellular recordings**

Mice were implanted with multi-electrode arrays for *in vivo* recordings of neural activity during awake behavior<sup>50, 59</sup>. Mice were implanted with two arrays, one targeting the OFC, and the other targeting the DMS and DLS. The array used in the OFC consisted of two-rows of eight platinum-plated tungsten electrodes (35  $\mu$ m, CD Neural), with electrodes spaced 150  $\mu$ m apart, and rows 200  $\mu$ m apart. For the OFC, arrays were centered A 2.6 mm and L 1.75 mm to Bregma, and V 2.25 to 2.4 mm from the surface of the brain. For the dorsal

striatum, the array consisted of two rows of eight electrodes (platinum-coated tungsten, 50  $\mu\text{m}$ , CD Neural), with electrodes spaced 200  $\mu\text{m}$  apart and row spacing of 1250  $\mu\text{m}$  so that one row targeted the DMS and the other row targeted the DLS. For the dorsal striatum, arrays were centered A0.5 mm and L1.75 mm and V 2.2 to 2.4 mm from the surface of the brain. Mice were perfused and brains fixed with 4% w/v paraformaldehyde, and array placement was verified using Nissl-stained brain slices (50  $\mu\text{m}$ ).

### Neural recordings during behavior

Mice were allowed at least 2 weeks of recovery before the start of behavioral and recording procedures. In brief, spike activity was recorded using the MAP system (Plexon Inc., TX) and initially sorted using an online-sorting algorithm. Mice were moved from one context to the other without disconnecting the head stage, and the same online sorting algorithm was used in both contexts on the same day. TTL pulses were used to synchronize the recordings with the lever-press behavior, to behaviorally timestamp the neural activity (10 ms resolution of the behavior). Data was then resorted offline (Offline Sorter, Plexon, Inc.) to identify single unit neuronal activity based on waveform, amplitude, and inter spike interval histogram (no spikes during a refractory period of 1.3ms)<sup>41</sup>. For dorsal striatum, in order to have mainly putative striatal medium spiny neurons in our analyses, units with a waveform trough half-width of less than 100  $\mu\text{s}$  and baseline firing rate more than 10 Hz, as well as those with a waveform trough half-width more than 250  $\mu\text{s}$  were excluded<sup>60</sup>. In OFC, units clustered around an amplitude of 150  $\mu\text{V}$ , waveform trough half-width of 200  $\mu\text{s}$ , and frequency of 3.5 Hz; in order to have mainly potential pyramidal neurons in our analyses, units with values 2 standard deviations greater than the population mean were excluded from the analyses.

### Lever-press related neurons

To examine task-related neural activity, for each previously isolated recorded unit we constructed a peri-event histogram (PETH) around time-stamped lever-press and head entry events, where neural activity was averaged in 20-ms bins, shifted by 1 ms and averaged across trials to analyze amplitude and latency during the recorded behaviors. Using the distribution of the PETH from 5000 to 2000 ms before the task as baseline activity, we slid 1 ms steps across 20-ms bins from 2000 ms before to 2000 ms after task-related events. A task-related neuron was up-modulated if it had a significant increase in firing rate defined as at least 20 consecutive overlapping bins with a firing rate larger than a threshold of 99% above baseline activity. An task-related neuron was down-modulated if it had a significant decrease in firing rate if at least 20 consecutive bins had a firing rate smaller than a threshold of 95% below baseline activity<sup>49,50</sup>. The onset of task-related activity was defined as the first of these 20 consecutive significant bins. Schedule-specific neurons were units that only showed a significant up—or down-modulation in the PETH around the behavioral event in the RI or RR context. Both-specific neurons were units that showed a significant up—or down-modulation in the PETH around the behavioral event in both RI and RR contexts. Rate modulation was defined as max or min firing rate in the time window from the beginning to the last of the consecutive significant bins minus baseline. The same analyses performed using a less conservative window of 1000 ms before and after task-events did not alter the present findings. See example average frequency traces (Fig 2c, g, and k).

## Statistical analyses

The  $\alpha$  level was set at 0.05 for all analyses performed except otherwise indicated. Initial analyses showed normal distributions for all behavioral data. All behavioral analyses and *in vivo* rate modulation data were analyzed using paired and unpaired t-tests, as well as 2-way and repeated measure ANOVAs with post-hoc analyses performed using Bonferroni-corrected paired t-tests where appropriate, including on normalized lever-presses during outcome reevaluation [normalization: (lever presses for Valued or Devalued states/Total lever presses Valued + Devalued states)]. We also included one-sample t-tests for normalized data to examine whether each condition differed from chance (0.5). That is, normalized data produced a distribution of lever-presses between Valued and Devalued states for each schedule, and value of 0.5 reflects the same level of lever-pressing between Valued and Devalued states. Chi-square ( $\chi^2$ ) analyses were used to look at proportional differences in percentage of lever-press related activity, direction of modulation, and the contributions of Both versus Specific neurons to the above changes. Correlation analyses were performed using Pearsons (r) correlation coefficient  $\alpha = 0.05$  for all tests performed.

Rate modulation values of lever-press related activity were used to calculate the modulation index for each neuron [(RR rate modulation-RI rate modulation)/(RR rate modulation + RI rate modulation)].

To investigate the shift in ensemble neural activity for each area in Fig. 5a and 5b, we calculated the difference between devalued and valued days in average rate modulation z-score around the lever-press for all lever-press related neurons (Both and Specific) within an area for each subject in RI and RR contexts.

To examine the degree of goal-directedness during outcome reevaluation (Fig. 3, 4, 5), we calculated a reevaluation index [(lever presses valued state - lever presses devalued state)/(lever presses valued state + lever presses devalued state)] for each mouse for the RR and RI contexts.

Correlation analyses were performed using Pearsons (r) correlation coefficient  $\alpha = 0.05$  for all tests performed. Data analyses were performed using Neuroexplorer, Graphpad Prism, and Matlab (Mathworks).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank David Lovinger, Eduardo Dias-Ferreira, Xin Jin, and Nicholas Oesch for comments on the manuscript. The DREADD virus was provided by the University of North Carolina Vector Core and Dr. R. Jude Samulski. This research was supported by the NIAAA Division of Intramural Clinical and Biological Research and European Research Council Grant (243393) and HHMI International Early Career Scientist Grant to R.M.C.

## References

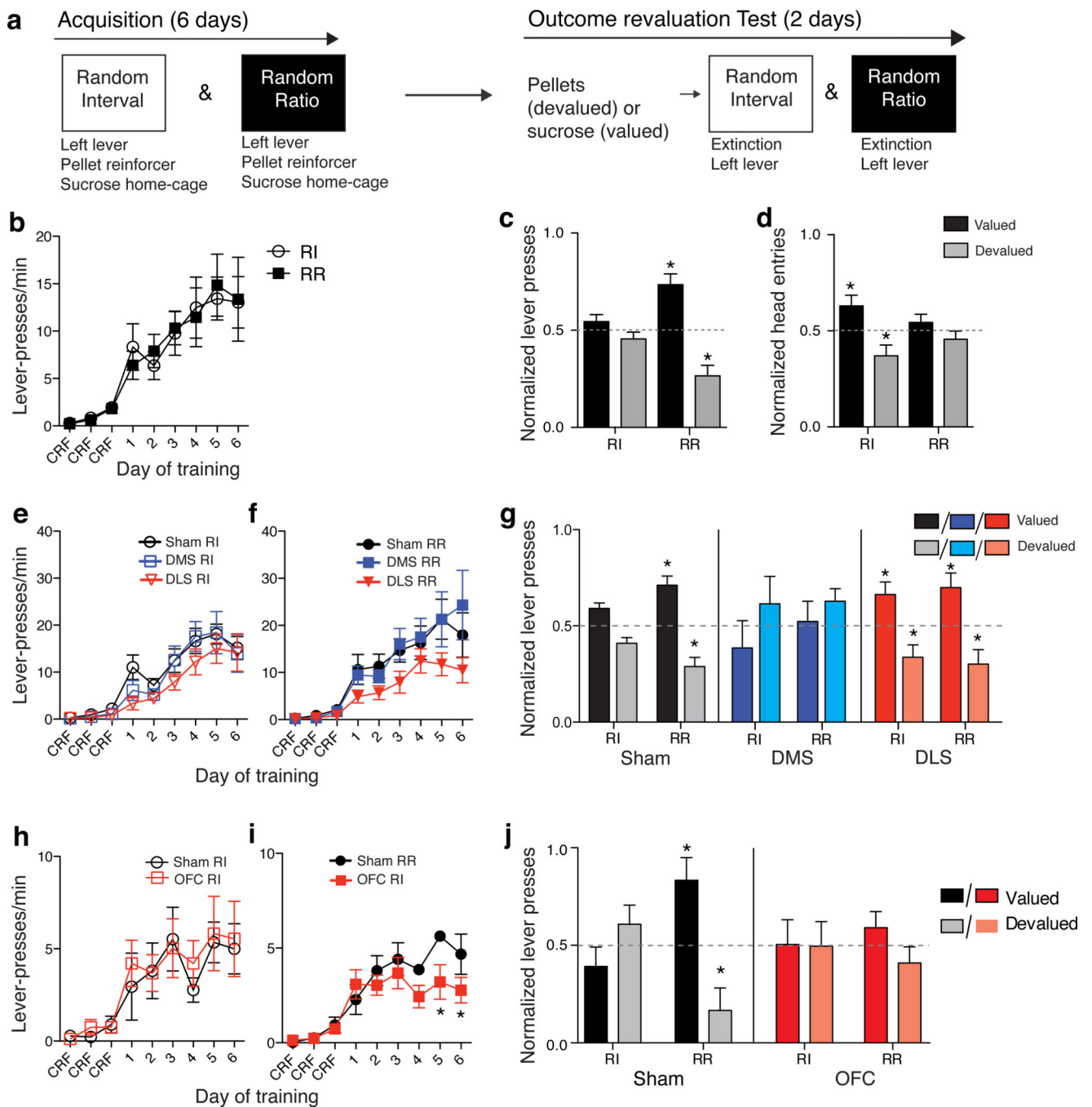
1. Adams CD. Variations in the sensitivity of instrumental responding to reinforcer devaluation. 1982

2. Adams C, Dickinson A. Instrumental Responding Following Reinforcer Devaluation. *Q J Exp Psychol-B*. 1981; 33:109–121.
3. Dickinson A. Actions and Habits: The Development of Behavioural Autonomy. *Philos Trans R Soc Lond. B, Biol Sci*. 1985; 308:67–78.
4. Colwill RM, Rescorla RA. Postconditioning devaluation of a reinforcer affects instrumental responding. *J Exp Psychol Anim Behav Process*. 1985; 11:120–132.
5. Dias-Ferreira E, et al. Chronic Stress Causes Frontostriatal Reorganization and Affects Decision-Making. *Science*. 2009; 325:621–625. [PubMed: 19644122]
6. Balleine BW, Delgado MR, Hikosaka O. The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience*. 2007; 27:8161–8165. [PubMed: 17670959]
7. Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci*. 2005; 8:1481–1489. [PubMed: 16251991]
8. Gillan CM, et al. Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am J Psychiatry*. 2011; 168:718–726. [PubMed: 21572165]
9. Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*. 2006; 7:464–476. [PubMed: 16715055]
10. Balleine BW, O'Doherty JP. Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology*. 2009; 35:48–69. [PubMed: 19776734]
11. Liljeholm M, O'Doherty JP. Contributions of the striatum to learning, motivation, and performance: an associative account. *Trends Cogn Sci (Regul Ed)*. 2012; 16:467–475. [PubMed: 22890090]
12. Pan WX, Mao T, Dudman JT. Frontiers: Inputs to the Dorsal Striatum of the Mouse Reflect the Parallel Circuit Architecture of the Forebrain. *Front Neuroanat*. 2010; 4:147. [PubMed: 21212837]
13. Voorn P, Vanderschuren LJMJ, Groenewegen HJ, Robbins TW, Pennartz CMA. Putting a spin on the dorsal-ventral divide of the striatum. *Trends Neurosci*. 2004; 27:468–474. [PubMed: 15271494]
14. McGeorge AJ, Faull RL. The organization of the projection from the cerebral cortex to the striatum in the rat. *Neuroscience*. 1989; 29:503–537. [PubMed: 2472578]
15. Yin HH, Ostlund SB, Knowlton BJ, Balleine BW. The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci*. 2005; 22:513–523. [PubMed: 16045504]
16. Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci*. 2004; 19:181–189. [PubMed: 14750976]
17. Yin HH, Knowlton BJ, Balleine BW. Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav Brain Res*. 2006; 166:189–196. [PubMed: 16153716]
18. Joel D, Doljansky J, Roz N, Rehavi M. Role of the orbital cortex and of the serotonergic system in a rat model of obsessive compulsive disorder. *Neuroscience*. 2005; 130:25–36. [PubMed: 15561422]
19. Rotge J-Y, et al. Meta-Analysis of Brain Volume Changes in Obsessive-Compulsive Disorder. *Biological Psychiatry*. 2009; 65:75–83. [PubMed: 18718575]
20. Atmaca M, et al. Volumetric MRI assessment of brain regions in patients with refractory obsessive-compulsive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2006; 30:1051–1057. [PubMed: 16687198]
21. Dickinson A, Balleine B. Motivational control of goal-directed action. *Animal Learning & Behavior*. 1994; 22:1–18.
22. Derusso AL, et al. Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. *Frontiers in integrative neuroscience*. 2010; 4
23. Balleine BW, Ostlund SB. Still at the choice-point: action selection and initiation in instrumental conditioning. *Ann N Y Acad Sci*. 2007; 1104:147–171. [PubMed: 17360797]
24. Corbit LH, Janak PH. Posterior dorsomedial striatum is critical for both selective instrumental and Pavlovian reward learning. *Eur J Neurosci*. 2010; 31:1312–1321. [PubMed: 20345912]



25. Walton ME, Behrens TEJ, Buckley MJ, Rudebeck PH, Rushworth MFS. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron*. 2010; 65:927–939. [PubMed: 20346766]
26. Gottfried JA, O'doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*. 2003; 301:1104–1107. [PubMed: 12934011]
27. Valentin VV, Dickinson A, O'Doherty JP. Determining the neural substrates of goal-directed learning in the human brain. *Journal of Neuroscience*. 2007; 27:4019–4026. [PubMed: 17428979]
28. Tanaka SC, Balleine BW, O'Doherty JP. Calculating consequences: brain systems that encode the causal effects of actions. *Journal of Neuroscience*. 2008; 28:6750–6755. [PubMed: 18579749]
29. O'Doherty JP. Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. *Ann N Y Acad Sci*. 2007; 1121:254–272. [PubMed: 17872386]
30. Schoenbaum G, Chiba AA, Gallagher M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat Neurosci*. 1998; 1:155–159. [PubMed: 10195132]
31. Izquierdo A, Suda R, Murray E. Bilateral Orbital Prefrontal Cortex Lesions in Rhesus Monkeys Disrupt Choices Guided by Both Reward Value and Reward Contingency. *Journal of Neuroscience*. 2004; 24:7540. [PubMed: 15329401]
32. Rudebeck P, et al. Frontal Cortex Subregions Play Distinct Roles in Choices between Actions and Stimuli. *Journal of Neuroscience*. 2008; 28:13775. [PubMed: 19091968]
33. Pickens C, et al. Different Roles for Orbitofrontal Cortex and Basolateral Amygdala in a Reinforcer Devaluation Task. *Journal of Neuroscience*. 2003; 23:11078. [PubMed: 14657165]
34. Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. *Nature*. 2006; 441:223–226. *Published online: 18 June 2008*; |. [PubMed: 16633341]
35. Plassmann H, O'doherty J, Rangel A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *Journal of Neuroscience*. 2007; 27:9984–9988. [PubMed: 17855612]
36. Young JJ, Shapiro ML. Dynamic coding of goal-directed paths by orbital prefrontal cortex. *Journal of Neuroscience*. 2011; 31:5989–6000. [PubMed: 21508224]
37. Takahashi YK, et al. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nat Neurosci*. 2011; 14:1590–1597. [PubMed: 22037501]
38. Kennerley SW, Behrens TEJ, Wallis JD. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci*. 2011; 14:1581–1589. [PubMed: 22037498]
39. Sul JH, Kim H, Huh N, Lee D, Jung MW. ScienceDirect.com - Neuron - Distinct Roles of Rodent Orbitofrontal and Medial Prefrontal Cortex in Decision Making. *Neuron*. 2010; 66:449–460. [PubMed: 20471357]
40. Hoover WB, Vertes RP. Projections of the medial orbital and ventral orbital cortex in the rat. *J Comp Neurol*. 2011; 519:3766–3801. [PubMed: 21800317]
41. Schilman EA, Uylings HBM, Galis-de Graaf Y, Joel D, Groenewegen HJ. The orbital cortex in rats topographically projects to central parts of the caudate-putamen complex. *Neurosci Lett*. 2008; 432:40–45. [PubMed: 18248891]
42. Corbit LH, Balleine BW. Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *Journal of Neuroscience*. 2005; 25:962–970. [PubMed: 15673677]
43. Wassum KM, Cely IC, Balleine BW, Maidment NT. Micro-opioid receptor activation in the basolateral amygdala mediates the learning of increases but not decreases in the incentive value of a food reward. *Journal of Neuroscience*. 2011; 31:1591–1599. [PubMed: 21289167]
44. Reynolds JN, Hyland BI, Wickens JRA. cellular mechanism of reward-related learning. *Nature*. 2001; 413:67–70. *Published online: 18 June 2008*; |. [PubMed: 11544526]
45. Gourley SL, Lee AS, Howell JL, Pittenger C, Taylor JR. Dissociable regulation of instrumental action within mouse prefrontal cortex. *Eur J Neurosci*. 2010; 32:1726–1734. [PubMed: 21044173]
46. Hilário MR. Endocannabinoid signaling is critical for habit formation. *Frontiers in integrative neuroscience*. 2007; 1
47. Ostlund SB, Balleine BW. Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *J Neurosci*. 2007; 27:4819–4825. [PubMed: 17475789]

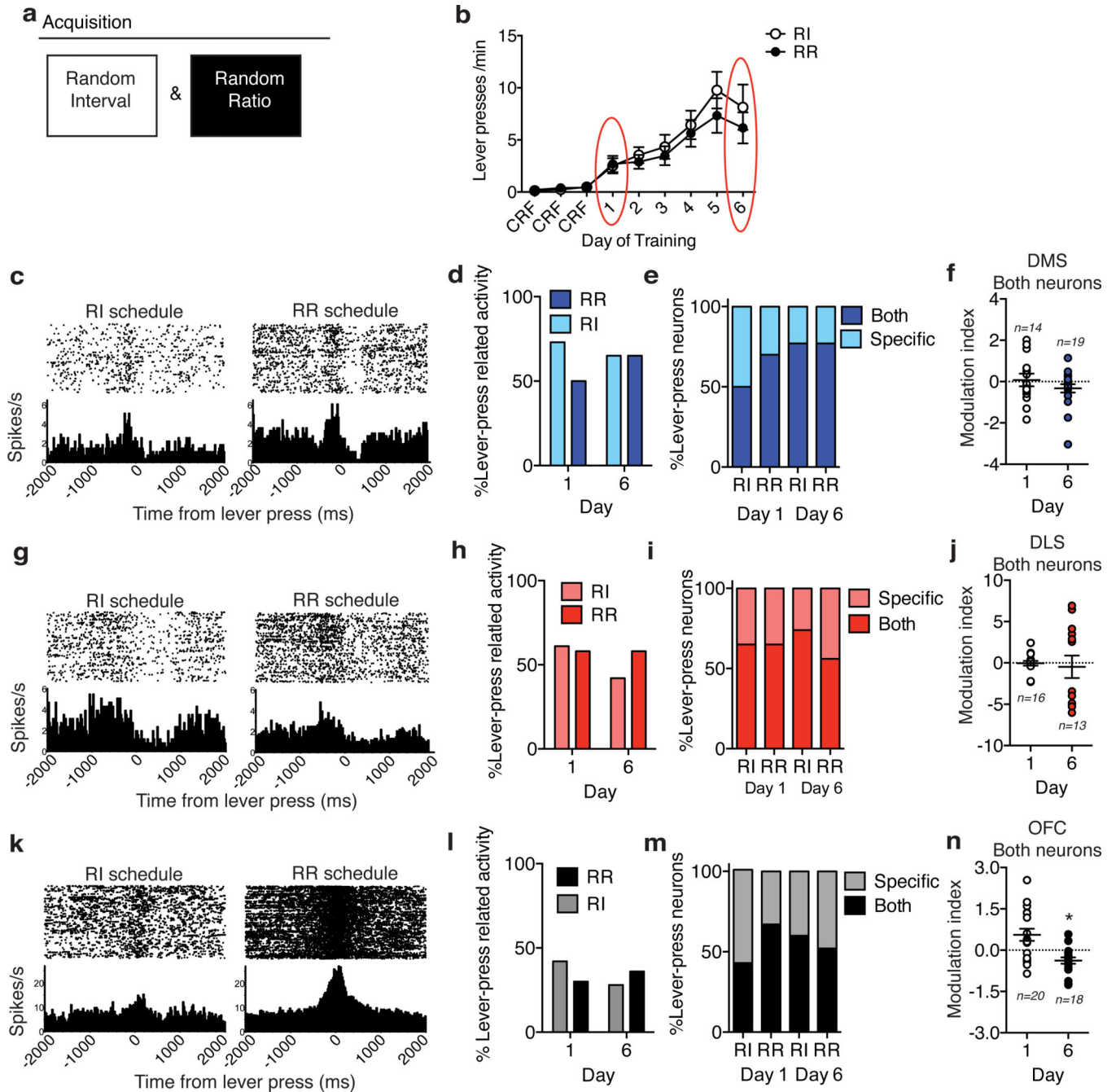
48. West EA, DesJardin JT, Gale K, Malkova L. Transient inactivation of orbitofrontal cortex blocks reinforcer devaluation in macaques. *Journal of Neuroscience*. 2011; 31:15128–15135. [PubMed: 22016546]
49. Belova MA, Paton JJ, Morrison SE, Salzman CD. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron*. 2007; 55:970–984. [PubMed: 17880899]
50. Jin X, Costa RM. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature*. 2010; 466:457–462. *Published online: 18 June 2008*; |. [PubMed: 20651684]
51. Thorn CA, Atallah H, Howe M, Graybiel AM. Differential Dynamics of Activity Changes in Dorsolateral and Dorsomedial Striatal Loops during Learning. *Neuron*. 2010; 66:781–795. [PubMed: 20547134]
52. Stalnaker TA, Calhoun GG, Ogawa M, Roesch MR, Schoenbaum G. Frontiers: Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. *Frontiers in integrative neuroscience*. 2010; 4:12. [PubMed: 20508747]
53. Dong S, Rogan SC, Roth BL. Directed molecular evolution of DREADDs: a generic approach to creating next-generation RASSLs. *Nat Protoc*. 2010; 5:561–573. [PubMed: 20203671]
54. Alexander GM, et al. Remote Control of Neuronal Activity in Transgenic Mice Expressing Evolved G Protein-Coupled Receptors. *Neuron*. 2009; 63:27–39. [PubMed: 19607790]
55. Tye KM, et al. Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature*. 2011; 471:358–362. *Published online: 18 June 2008*; |. [PubMed: 21389985]
56. Rushworth MFS, Noonan MP, Boorman ED, Walton ME, Behrens TE. Frontal cortex and reward-guided learning and decision-making. *Neuron*. 2011; 70:1054–1069. [PubMed: 21689594]
57. Colwill, RM.; Rescorla, RA. The psychology of learning and motivation. Bower, G., editor. New York: academic; 1986. p. 55-104.
58. Remijne PL, et al. Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. *Archives of General Psychiatry*. 2006; 63:1225–1236. [PubMed: 17088503]
59. Costa RM, Cohen D, Nicolelis MAL. Differential corticostriatal plasticity during fast and slow motor skill learning in mice. *Curr Biol*. 2004; 14:1124–1134. [PubMed: 15242609]
60. Burkhardt JM, Jin X, Costa RM. Dissociable effects of dopamine on neuronal firing rate and synchrony in the dorsal striatum. *Frontiers in integrative neuroscience*. 2009; 3:28. [PubMed: 19949467]



**Figure 1. Shifting between goal-directed and habitual actions**

(a) Schematic of the within-subject behavioral design. Each day mice were trained to press the same RI manipulandum (identical lever in same position) for the same reinforcer on a RI schedule in one context, and on a RR schedule in a separate context. A control reinforcer was presented later in their home cage. After acquisition, mice were given a sensory-specific outcome revaluation test where they could free-feed on either the control (valued state) or the previously earned reinforcer (devalued state). Mice were then immediately placed into one followed by the other training context for 5 minutes and non-reinforced lever-presses

were measured. **(b)** Average lever-presses/min across CRF and concurrent RI and RR schedule training for C57BL/6J mice ( $n = 10$ ). **(c, d)** Subsequent average lever presses **(c)** and head entries **(d)** (normalized to Revaluation state) made in RI and RR training contexts in Valued and Devalued states. **(e, f, h, i)** Average lever-presses/min across CRF and concurrent RI and RR schedule training for Sham ( $n = 9$ ), DMS ( $n = 5$ ), and DLS ( $n = 7$ )lesioned mice **(e, f)** and for Sham ( $n = 7$ ) and OFC ( $n = 5$ )lesioned mice **(h, i)**. **(g, j)** Average normalized to Revaluation state lever presses made in RI and RR training contexts during Valued and Devalued states for Sham, DMS, and DLS lesioned mice, **(g)** and Sham and OFC lesioned mice **(j)**. Repeated measures ANOVA and one-way sample t-tests were used. Error bars indicate s.e.m. \* =  $p < 0.05$ .  $n$ 's = 5–10 per group. (See also Supplementary Fig. S1–4).



**Figure 2. Action-encoding in different corticostriatal loops during RI and RR schedule training** (a) Schematic of the within-subject behavioral acquisition design and (b) rate of lever pressing under RI and RR schedules for recording mice ( $n = 8$ ). Example raster plots and peri-event time histograms (PETH) of the same DMS (c), DLS (g), and OFC (k) neuron showing lever-press related activity under RI and RR reinforcement schedules on Day 6 of training. Each row in the raster is neural activity  $\pm 2$  s around a lever press (time = 0). Trials are sorted according to the order of lever-presses made across the session. The percentage of lever-press related activity per mouse during RI and RR schedule acquisition for DMS (d),

DLS (**h**), and OFC (**l**). The percentage of lever-press related neurons per mouse that change firing rate during lever-press behavior in both RI and RR (Both-schedule neurons) or only during lever-press behavior in RI or RR (Specific) in DMS (**e**), DLS (**i**), and OFC (**m**) across RI and RR acquisition. The modulation index for Both-schedule neurons across acquisition in DMS (**f**), DLS (**j**), and OFC (**n**).  $\chi^2$  analyses, unpaired t-tests, and one-sample t-tests were used. Error bars indicate s.e.m. \* =  $p < 0.05$ . (See also Supplementary Fig. S5–7).

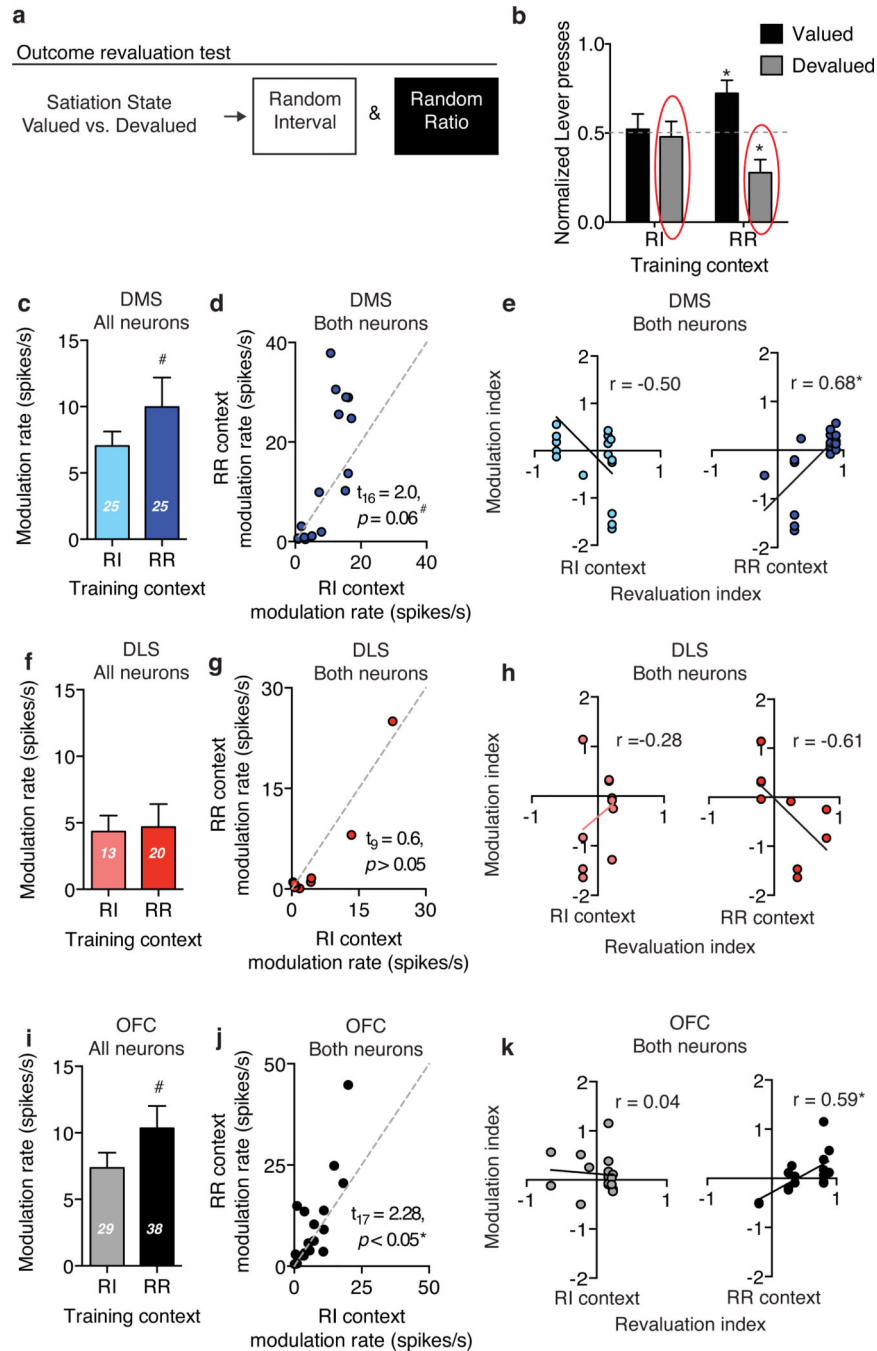
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Figure 3. Devalued state encoding of goal-directed and habitual actions in corticostriatal circuits** (a) Schematic of the within-subject outcome revaluation testing, and (b) normalized lever-pressing on Valued and Devalued days in previously RI and RR trained contexts. Modulation rate (absolute change in firing rate) of lever-press related DMS (c), DLS (f), and OFC (i) neurons (number in bar graph =  $n$  of modulated recorded neurons) during lever-press behavior in previously trained RI and RR contexts in the Devalued state. X-Y scatter-plots of Both-schedule neuron modulation during lever-press behavior in RI vs. RR contexts in DMS (d), DLS (g), and OFC (j) in the Devalued state. Correlations between the modulation

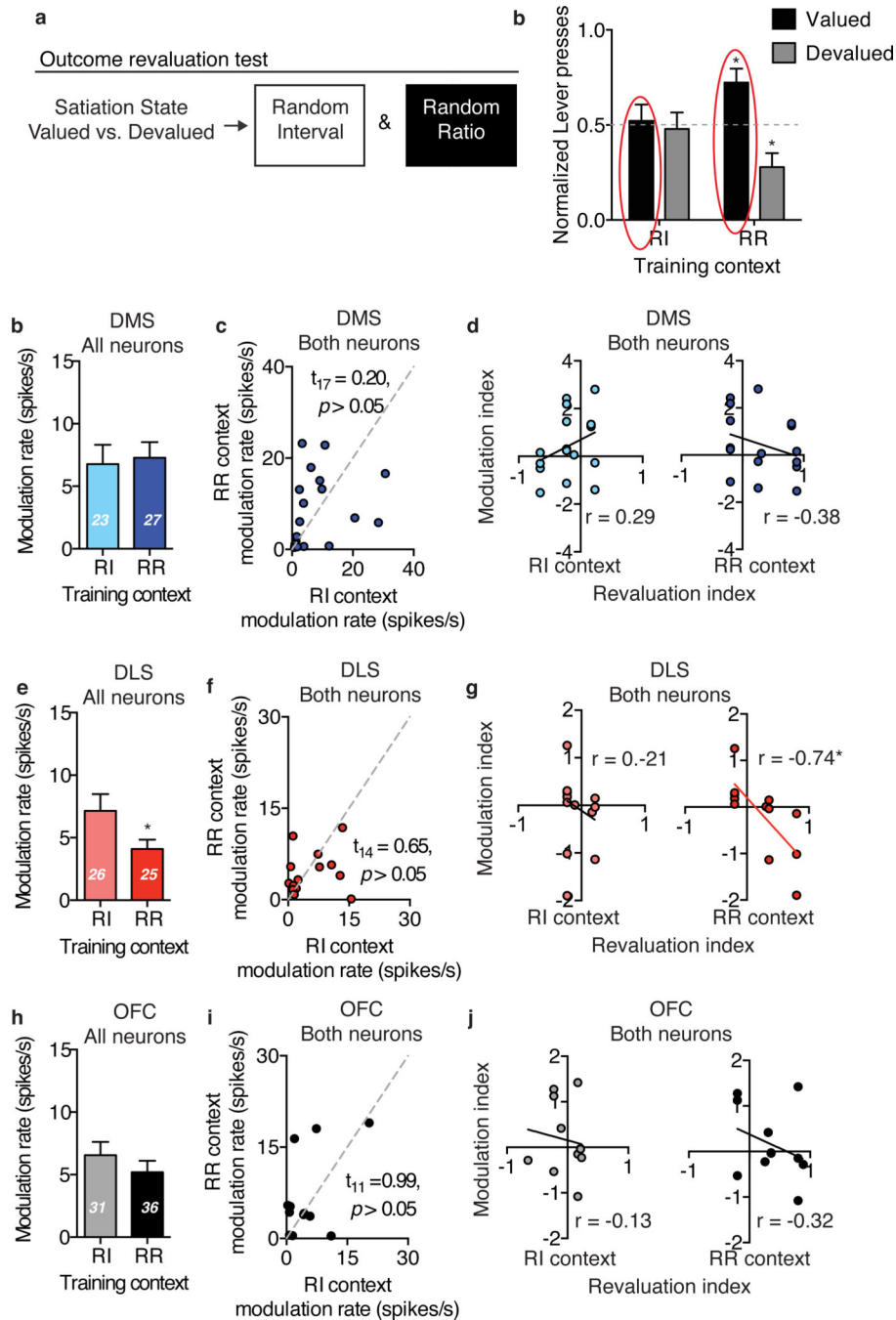
index of Both-schedule neurons in the Devalued state and the revaluation index for mice in RI and RR contexts for DMS (**e**), DLS (**h**), and OFC (**k**) Both-schedule neurons. Repeated-measures ANOVA, one-sample, unpaired, and paired t-tests, and Pearson correlation analyses were used. Error bars indicate s.e.m. \* =  $p < 0.05$ . (See also Supplementary Figs. S8 and S9).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 4. Valued state encoding of goal-directed and habitual actions in corticostriatal circuits** (a) Schematic of the within-subject outcome revaluation testing, and (b) normalized lever-pressing on Valued and Devalued days in previously RI and RR trained contexts. Modulation rate (absolute change in firing rate) of lever-press related DMS (c), DLS (f), and OFC (i) neurons (number in bar graph =  $n$  of recorded modulated neurons) during lever-press behavior in previously trained RI and RR contexts in the Valued state. X-Y scatter-plots of Both-schedule neuron modulation during lever-press behavior in RI vs. RR contexts in DMS (d), DLS (g), and OFC (j) in the Valued state. Correlations between the modulation

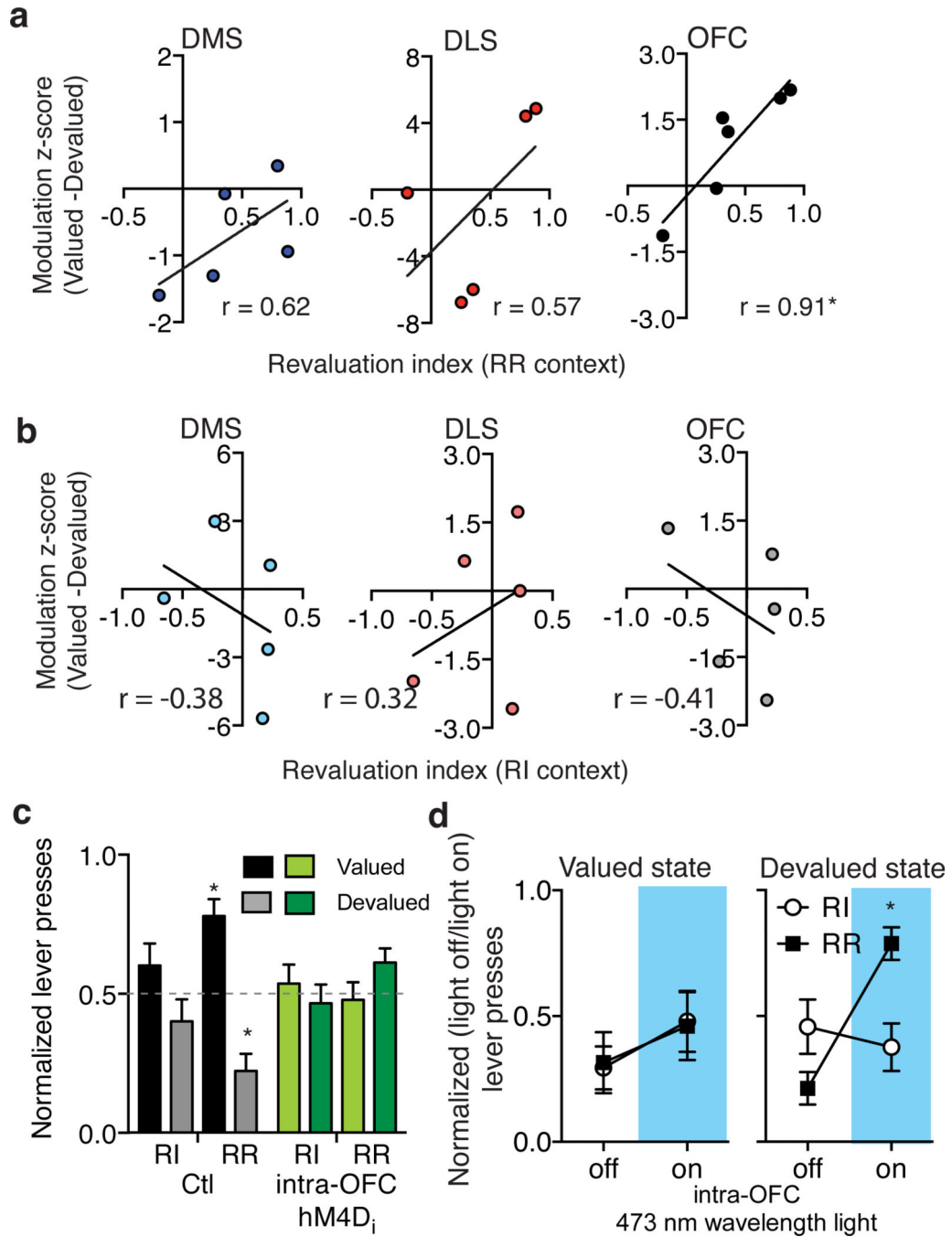
index of Both-schedule neurons in the Valued state and the revaluation index for mice in RI and RR contexts for DMS (**e**), DLS (**h**), and OFC (**k**) Both-schedule neurons. Repeated-measures ANOVA, one-sample, unpaired, and paired t-tests, and Pearson correlation analyses were used. Error bars indicate s.e.m. \* =  $p < 0.05$ . (See also Supplementary Figs. S8 and S9).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 5. OFC conveys information about changes in action value**

Shift in OFC, DMS, and DLS neural ensemble modulation between valued and devalued states for each mouse (changes in Z-scores of lever-press related modulation for Both-schedule and Schedule specific neurons), correlated with the magnitude of goal-directed(a) and habitual behavior (b) in the same animal as measured by a Revaluation index. (c) Effect of chemogenetic inhibition of OFC projection neurons on lever-press (normalized to Revaluation state) behavior during outcome revaluation testing. Following either an OFC bilateral co-injection of cre-dependent hM4D<sub>i</sub> receptors and cre-recombinase expressed

under the *CAMKII $\alpha$*  promoter (intra-OFC hM4D<sub>i</sub>;  $n = 10$ ) or cre-dependent hM4D<sub>i</sub> and a GFP virus (Ctl;  $n = 11$ ), mice were trained concurrently on RI and RR schedules using the within-subject design. On Valued and Devalued days, mice were given a systemic 1 h pretreatment with clozapine-*N*-oxide (CNO) (1mg/kg), and subsequent lever-press behavior was recorded in each context. **(d)** Effect of bilateral optogenetic activation of OFC on lever-press (normalized to light on/off for each Revaluation state) behavior during outcome revaluation testing. Following bilateral-OFC injection of ChR2-YFP expressed under the *CAMKII $\alpha$*  promoter and implantation of bilateral optic fiber ferrules, mice ( $n = 6$ ) were trained concurrently on RI (open circles) and RR (black squares) schedules using the within-subject design. On Valued and Devalued days, lever-press behavior was recorded in each context for an initial 5 min without photostimulation, and during subsequent 5 min photostimulation with 10 hz, 5 ms pulses of 473 nm wavelength light. Repeated-measures ANOVA, one-sample, paired t-tests, and Pearson correlation analyses were used. Error bars indicate s.e.m. \* =  $p < 0.05$ . (See also Supplementary Figs. S10-S12).