Choice suppression through opponent but not independent function of the striatal indirect pathway

Kristen Delevich[1], Benjamin Hoshal[3], Anne GE Collins[1], Linda Wilbrecht[1,2,4]*

1 Department of Psychology, University of California, Berkeley, Berkeley, CA 94720, USA
2 Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94720 USA
3 Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley CA 94720 USA
4 Lead Contact

* corresponding author: wilbrecht@berkeley.edu

Significance

There is significant clinical value to understanding how we reject or suppress making a choice, and the dorsomedial striatum (DMS) is a critical arbiter of this process. While optogenetic stimulation of DMS indirect pathway spiny projection neurons (iSPNs) can inhibit movement, it is unclear how iSPNs contribute to suppression of choices. A simple 'no go' function has been proposed for iSPNs, suggesting their activity enables choice suppression, but we found that chemogenetic activation of iSPNs impaired suppression of low value choices. This effect was explained by an algorithmic model in which the relative output of direct pathway (dSPNs) and iSPNs determines choice. Our findings have important implications for designing interventions to improve maladaptive decision-making in psychiatric disorders and addiction.

Abstract

The dorsomedial striatum (DMS) plays a key role in action selection, but little is known about how direct and indirect pathway spiny projection neurons (dSPNs and iSPNs) contribute to serial decision-making. A popular 'select/suppress' heuristic proposes that dSPNs encode selected actions while iSPNs encode the suppression of alternate actions. Here, we used pathway-specific chemogenetic manipulation during serial choice behavior to test predictions generated by the 'select/suppress' heuristic versus a network inspired OpAL (Opponent Actor Learning) model of basal ganglia function in which the relative balance of dSPN and iSPN output determines choice. In line with OpAL predictions, chemogenetic activation, not inhibition, of iSPNs

disrupted learned suppression of nonrewarded choices. These results cannot be explained by the classic view that choice suppression is an extension of iSPN stopping or 'no go' function. Together, our computational and empirical data challenge the 'select/suppress' interpretation of striatal function in the context of choice behavior and highlight the ability of iSPNs to modulate choice exploration.

Keywords: striatum, decision-making, reinforcement learning

Introduction

In everyday decision-making, we often select among options in a serial fashion, foregoing low value choices in order to arrive at a higher value choice. The inability to suppress poor choices is a core component of addiction, eating disorders, and obsessive-compulsive disorder (1-3). The mechanisms underlying choice suppression are therefore highly relevant to psychiatry and public health.

The DMS (homologous to the primate caudate) is a key brain structure for goal-directed action selection (4-8), and striatal dysfunction is associated with maladaptive choice behavior (9-13). However, it is still poorly understood how choice selection and suppression are implemented at the circuit level (14). Furthermore, much of the relevant functional data comes from two-alternative forced choice (2AFC) tasks (15-19), in which it is difficult to dissociate the selection of one choice (e.g. turn left) from the suppression of another (e.g. do not turn right). Therefore, studying DMS function in the context of a serial task in which animals move freely and select among multiple options may reveal new insights into the circuit mechanisms that underlie choice selection and suppression.

The DMS is primarily composed of D1 receptor expressing dSPNs and D2 receptor expressing iSPNs (20), whose activity reflect task features including movement, cues, and value (16, 21-26). Consistent with predictions from functional neuroanatomy (27-29) and theoretical work (30, 31), optogenetic stimulation of dSPNs promotes movement and reinforces actions ('go' functions) (32-34) whereas optogenetic stimulation of iSPNs inhibits movement and drives aversion ('no go' functions) (32-34). In a 2AFC task, dSPN stimulation promotes contraversive choices whereas iSPN stimulation promotes ipsiversive choices in a manner that is reward history dependent (18). While these data suggest that the function of dSPNs and iSPNs in decision-making are dichotomous, it is important to note that they are co-active during goal-directed movement (23, 35, 36). It has been suggested that the two pathways work in concert such that dSPNs promote desired actions/choices whereas iSPNs suppress competing actions/choices

(27, 29, 35, 37), but this interpretation, which we refer to as the ‘select/suppress’ heuristic, has not been directly tested in the context of serial choice.

Our current study aimed to overcome this knowledge gap by testing a hypothesis that follows from the ‘select/suppress’ heuristic, that increasing iSPN activity should aid in the suppression of a choice whereas blocking iSPN activity should lead to a failure to suppress choice. To this end, we trained mice in an odor-guided serial choice task in which they approach multiple options before making a choice selection. Through trial and error, mice learn that only one of four odors is rewarded and learn to suppress choice to nonrewarded odors. After learning, we chemogenetically manipulated DMS dSPNs or iSPNs and examined choice behavior. To interpret our behavioral data, we compared predictions made by the ‘select/suppress’ heuristic and surprisingly opposite predictions that emerge from an algorithmic model of basal ganglia function, the Opponent Actor Learning (OpAL) model.

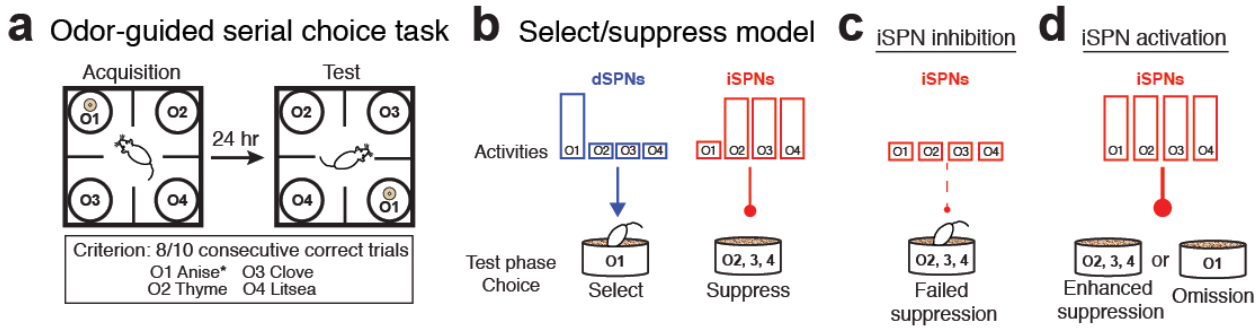
Results

In order to quantify selection and suppression of choices, we trained mice in an odor-guided serial choice task in which mice approach multiple distinctly scented pots in a serial fashion, rejecting pots until they choose one by digging in the scented shavings it contains (38) (Fig. 1A). Only one odor is rewarded (O1, “anise”), and the odor-action-reward contingency is learned through trial and error during an Acquisition phase. At the start of Acquisition, mice consistently exhibit an initial preference for a nonrewarded odor (O4, “thyme”). Therefore, in addition to learning to choose O1, a large part of Acquisition training is learning to suppress choice to O4. Twenty-four hours later mice enter a recall Test phase where their ability to select the rewarded odor (O1) and suppress choice to the remaining three nonrewarded odors (O2-4) is assessed (Fig. 1a, and Online Methods).

The ‘select/suppress’ model predicts inhibition of the indirect pathway should induce failure to suppress nonrewarded choices

If iSPNs are responsible for choice suppression as suggested by the ‘select/suppress’ heuristic model framework (Fig. 1B), then inhibition of iSPNs during the Test phase should lead to more errors, indicating a failure to suppress choice to nonrewarded pots (Fig. 1C). In this same framework, activation of iSPNs should facilitate suppression of nonrewarded choices and thus reduce errors and improve performance, or alternatively produce choice omission (Fig. 1D).

Fig. 1: Odor-guided serial choice task and ‘select/suppress’ heuristic predictions for iSPN manipulation



(**a**) 4 option odor-based serial choice task. (**b**) A ‘select/suppress’ model emphasizes the independent role of iSPNs in suppressing choices to nonrewarded odor options. (**c,d**) The ‘select/suppress’ heuristic predicts iSPN inhibition leads to failure to suppress low value choices while iSPN activation should enhance suppression of low value choices or increase omissions.

The OpAL model predicts that activation, not inhibition, of the indirect pathway should induce failure to suppress nonrewarded choices

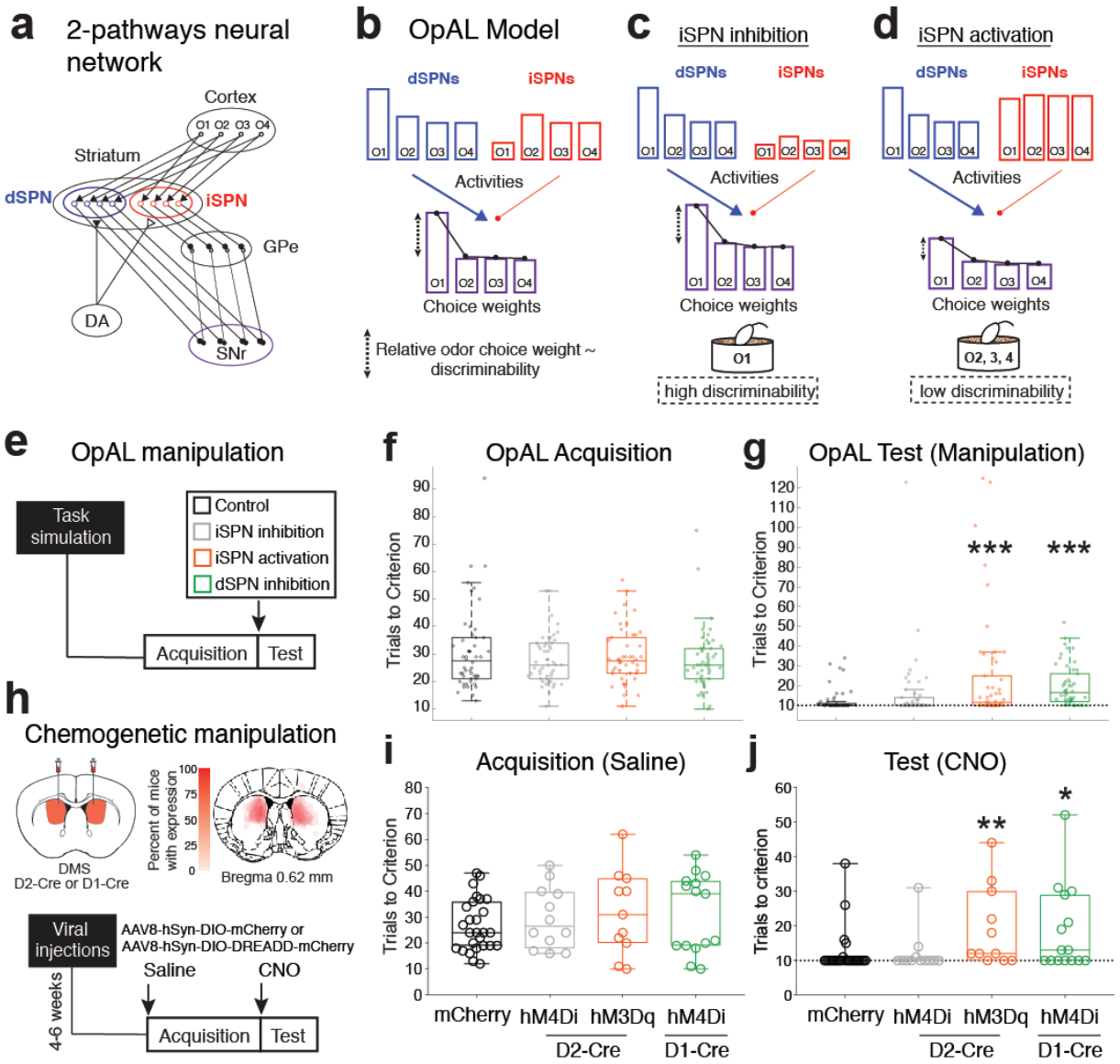
Alternate, network-inspired models of basal ganglia function (31) and recent *in vivo* data of dSPN and iSPN activity suggest that an independent division of labor does not properly account for how action and choice arise from the activity of dSPN and iSPN populations (24, 39-43). An alternate model that accounts for the opponent nature of the two pathways may generate different and more accurate predictions. Therefore, we turned to OpAL (Opponent Actor Learning), an algorithmic model of a biologically plausible basal ganglia network (Fig. 2A) in which choice is a function of the weighted difference between dSPN and iSPN population activity (Fig. 2B, Fig. S1; see Methods). OpAL predicted that decreasing iSPN activity would not alter the relative difference between choice weights, leaving discriminability among odors options high (Fig. 2C) but predicted that increasing iSPN activity would minimize the difference in choice weights between the rewarded odor and the nonrewarded odors, lowering discriminability (Fig. 2D). We simulated performance in the odor-guided serial choice task by adjusting iSPN population activity during the Test phase of the task via parameters that mimicked chemogenetic activation or inhibition (Fig. 2E). OpAL simulations predicted that activation of iSPNs would increase Test phase errors (i.e. choices to nonrewarded odors) while inhibition of iSPNs should not affect performance (Fig. 2F,G). The OpAL model therefore made predictions about iSPN function that were opposite to the predictions generated by the 'select/suppress' heuristic.

Chemogenetic manipulation experiments confirm activation, not inhibition, of the indirect pathway induces failure to suppress nonrewarded choices

To directly test the predictions made by the 'select/suppress' heuristic and OpAL model we turned to *in vivo* chemogenetic manipulation. D1-Cre or D2-Cre mice were injected bilaterally into the DMS with 0.5 µL of virus and 4-6 weeks later were trained in the 4 option odor-guided serial choice task (Fig. 2H). Mice that expressed Cre-inducible mCherry were used to control for any effects of surgery, AAV infection, and CNO administration on behavior. The efficacy of

activating and inhibitory DREADD manipulation was confirmed in slice electrophysiology experiments; CNO activation of hM4Di suppressed iSPN synaptic release and CNO activation of hM3Dq depolarized iSPNs (Fig. S2). Viral targeting was confirmed and mapped for all mice tested (Fig. S3). No differences in Acquisition learning, measured as trials to criterion (TTC), were observed across groups (Kruskal Wallis test, $H=1.42$, $p= 0.70$) (Fig. 2I). Twenty-four hours after Acquisition, mice were administered CNO (1.0 mg/kg, i.p.) and run in the recall Test phase, where we examined their selection of the rewarded scented pot (O1) and successful or unsuccessful suppression of the remaining three nonrewarded pot (O2-4) choices. In the Test phase, mCherry control and D2-Cre inhibitory DREADD (hM4Di) groups exhibited robust recall of the rewarded choice and successful suppression of the nonrewarded choices, with most mice reaching criterion in the minimum number of trials required (median TTC = 10, IQR = 0) (Fig. 2J). Meanwhile, mice expressing activating DREADD in iSPNs (D2-hM3Dq) and mice expressing inhibitory DREADD in dSPNs (D1-hM4Di) took significantly more trials to reach criterion compared to mCherry controls (Kruskal-Wallis test, $H=12.4$, $p= 0.006$; Dunn's uncorrected post hoc test, $^{**}p<0.01$ D2-hM3Dq vs. mCherry, $^{*}p<0.05$ D1-hM4Di vs. mCherry) (Fig. 2J). These data were consistent with the predictions made by the OpAL model (Fig. 2G).

Fig. 2: OpAL model and chemogenetic manipulation data show iSPN activation not inhibition impairs suppression of low value choices



(**a**) Schematic of cortico-basal ganglia network. The 4 odor options are represented as separate action channels in direct and indirect pathways. (**b**) OpAL model with example of choice among the 4 odor options in the serial choice task. Here, weights onto dSPNs and iSPNs for different odor stimuli reflect learned values updated by trial and error RL mechanism during the Acquisition phase. The weighted difference between dSPN activity and iSPN (Choice weights) are then transformed into choice probabilities using the softmax function. (**c**) OpAL predicts that iSPN inhibition increases choice weights across odor options but that the relative value difference (discriminability) between odor choice weights is maintained. (**d**) OpAL predicts that activation of iSPN minimizes the difference in choice weights across odor choices which would render choice exploratory. (**e**) OpAL simulated trial histories for the odor-guided serial choice task, with manipulation (control, inhibition, or activation) applied to dSPNs or iSPNs only during the Test phase of the task. (**f**) OpAL simulated trials to criterion during Acquisition phase (unmanipulated)

do not differ across groups (p>0.05, Kruskal Wallis test). (**g**) iSPN activation and dSPN inhibition during Test phase significantly increased trials to criterion in OpAL-simulated data compared to controls (***p<.0010, Kruskal Wallis test). OpAL simulation of iSPN inhibition during Test phase did not significantly affect performance (p= 0.11, Kruskal Wallis test). (**h**) Chemogenetic manipulation of DMS iSPNs or dSPNs in D2-Cre or D1-Cre mice. (**i**) All groups perform similarly in the Test phase (H=1.42, df=3, p=0.70, Kruskal Wallis test). (**j**) There was a significant effect of virus on performance during the Test phase (H=12.46, df=3, **p<0.01, Kruskal Wallis test), with D2-Cre mice expressing DIO-hM3Dq and D1-Cre mice expressing DIO-hM4Di taking significantly more trials to complete the Test phase (**p<0.001, mCherry vs. D2-hM3Dq, *p<0.05, mCherry vs. D1-hM4Di, post hoc uncorrected Dunn's multiple comparison test).

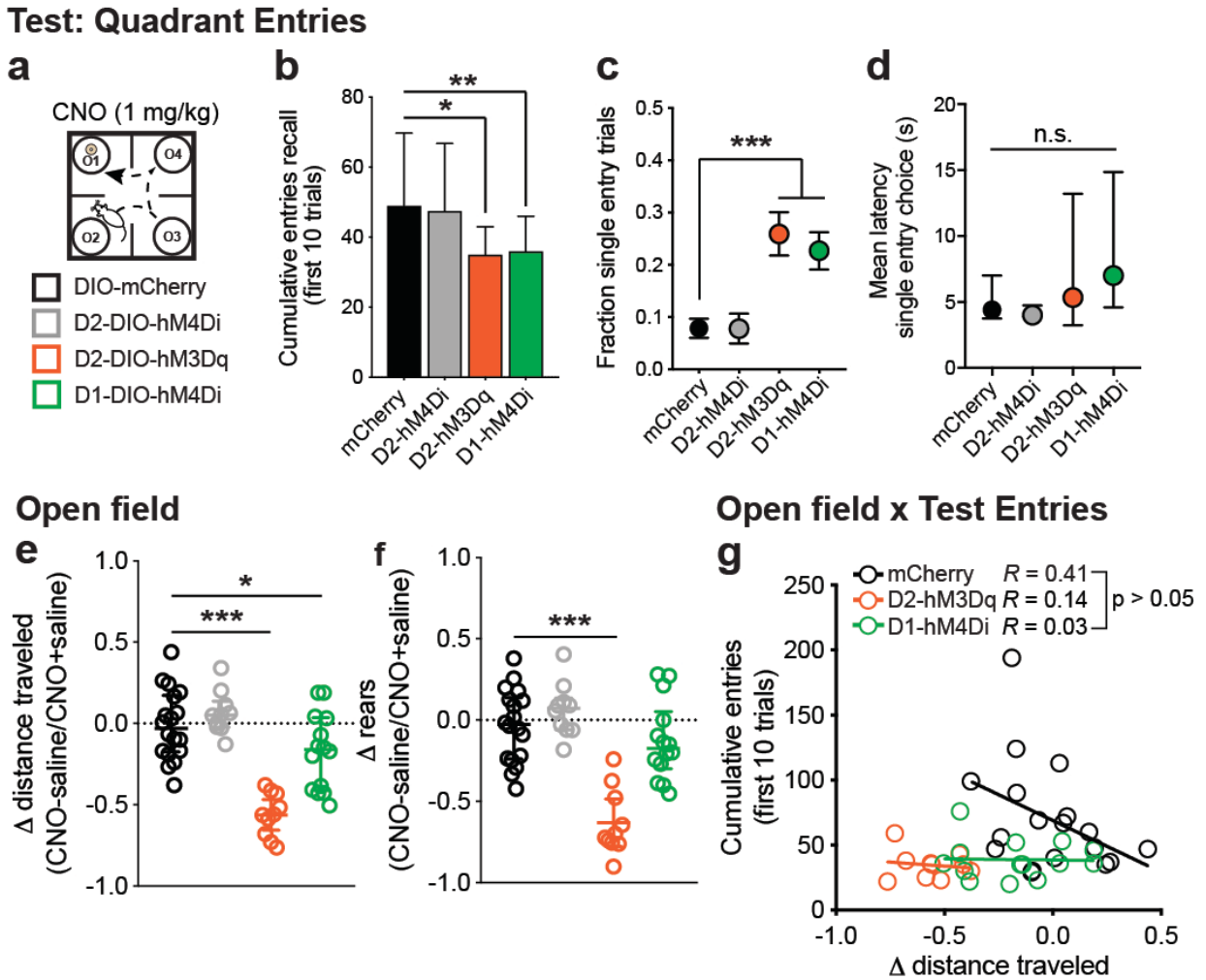
Acute chemogenetic manipulation alters choice strategy in a manner not explained by locomotor effects

To better understand the nature of the chemogenetic effect we next analyzed more fine-grained aspects of behavior in the serial choice task and other motor behaviors. During each trial of the odor-based serial choice task, mice were free to enter each of the 4 quadrants and sample the odors present in each pot, quantified as entries, before making a bi-manual dig to indicate their choice (Fig. 3A). Mice expressing activating DREADD in iSPNs (D2-hM3Dq) or inhibitory DREADD in dSPNs (D1-hM4Di) consistently made fewer entries during the Test phase compared to mCherry controls and mice expressing inhibitory DREADD in iSPNs (D2-hM4Di) (Fig. 3B). Both D2-hM3Dq and D1-hM4Di mice were more likely to choose the first odor they encountered (classified as single entry trials) compared to mCherry control and D2-hM4Di mice (Fig. 3C). We next asked whether this increase in single entry trials could be explained by impulsivity, a change in motivation, or movement ability. We found that the rank odor of choice preferences in Acquisition was intact during the Test phase in all groups, indicating that even on single entry trials, mice used odor value information to guide choice, inconsistent with impulsivity (Fig. S4). In addition, overall choice latency (Fig. S4) and single entry trial latency did not differ across groups (Fig. 3D). D2-hM3Dq and D1-hM4Di mice completed more trials (indicated by greater trials to criterion) and had similar numbers of omission trials compared to mCherry controls (Fig. S5) suggesting that they remained motivated to perform the task.

To examine if the reduction in entries during Test phase reflected changes in movement, we measured the effect of CNO on spontaneous locomotion and rotarod performance in all groups

at least one week after conclusion of the odor-guided serial choice task. CNO administration significantly reduced spontaneous locomotion in mice expressing activating DREADD in iSPNs (D2-hM3Dq) by ~50% compared to mCherry control mice (Fig. 3E). Mice expressing inhibitory DREADD in dSPNs (D1-hM4Di) showed a less dramatic reduction in distance traveled on CNO, whereas mice expressing inhibitory DREADD in iSPNs (D2-hM4Di) did not differ from mCherry control mice on CNO (Fig. 3E). However, the effect of CNO on spontaneous locomotion did not correlate with entries made during Test phase (Fig. 3G) suggesting that the influence of CNO on behavior in the Test phase is not simply motor-related. In addition, CNO did not alter rotarod performance in any groups tested (Fig. S6).

Fig. 3: Chemogenetic manipulation of direct and indirect pathway neurons alters odor option sampling in a manner dissociable from motor effects

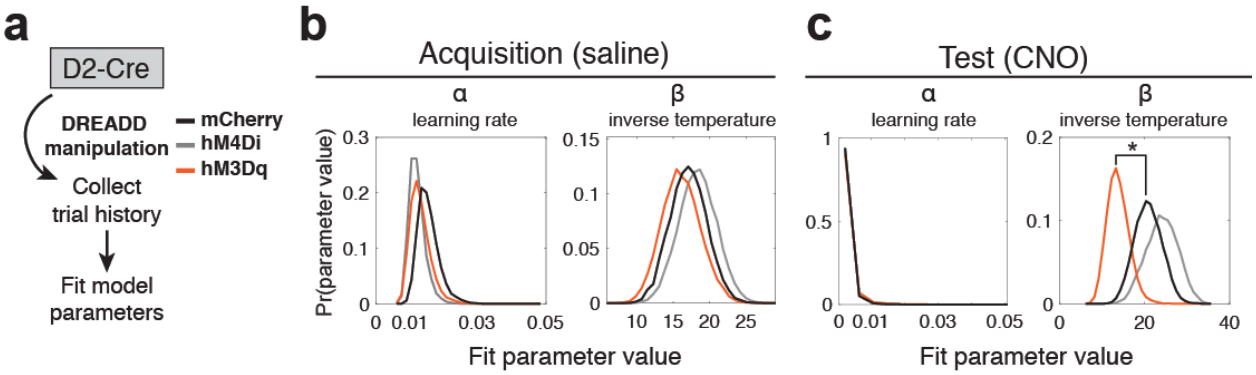


(**a**) Test phase quadrant entries. (**b**) D2-Cre mice expressing activating DREADD (D2-hM3Dq) or D1-Cre mice expressing inhibitory DREADD (D1-hM4Di) made fewer entries during Test phase on CNO compared to mCherry control mice on CNO (*p<0.05, **p<0.01 Kruskal Wallis ANOVA). (**c**) D2-hM3Dq and D1-hM4Di mice made significantly more choices to the odor in the first quadrant they entered, referred to as single entry trials ($F(3,40.65)= 9.55$, ***p< 0.0001 main effect of group Brown-Forsythe ANOVA; **p<0.01 unpaired t-test with Welch's correction). (**d**) Latency of single entry trials did not differ across groups. (**e**) Spontaneous locomotion was significantly reduced in D2-hM3Dq mice on CNO (1 mg/kg) compared to saline ($F(3,48)= 22.38$, ***p<0.0001, one-way ANOVA, ***p<0.0001, uncorrected Fisher's LSD) and D1-hM4Di mice to a lesser extent (*p<0.05, uncorrected Fisher's LSD). (**f**) CNO administration significantly reduced the number of vertical rears made by D2-hM3Dq mice ($F(3,48)= 21.87$ ***p<0.0001, one-way ANOVA, ***p<0.0001 mCherry vs. D2-hM3Dq uncorrected Fisher's LSD). (**g**) Locomotor modulation on CNO did not correlate with the number of entries mice made on CNO during Test phase for any group.

Trial-by-trial RL modeling suggests that enhancing iSPN activity alters Test phase performance by increasing choice stochasticity

Finally, we compared multiple reinforcement learning (RL) models (44) fit to trial-by-trial changes in behavior of using a hierarchical fitting process to determine whether the pattern of odor selection we observed was due to a change in choice policy (Fig. 4A and Online Methods). The best fit model for our behavioral data included phase specific parameters for the learning rate α and the inverse temperature parameter β, which captures choice stochasticity (see Table 1 for alternate model comparison). Focusing on D2-Cre mice, we found that Acquisition phase α and β parameters did not differ across groups (Fig. 4B), whereas the Test phase β parameter was significantly lower for mice expressing activating DREADD in iSPNs (D2-hM3Dq) compared to mCherry control and inhibitory DREADD (D2-hM4Di) (Fig. 4C). These data suggest that chemogenetic activation of iSPNs in the DMS makes choice policy more stochastic, and in the context of our task in which only one option is rewarded, a more exploratory choice policy leads to worse performance. Model fits of OpAL simulated trial histories converged on the same results, with iSPN activation associated with decreased Test phase β (Fig. S7). RL model fits to OpAL simulated data also suggested that inhibition of dSPNs reduce the Test phase β parameter, but RL fits to D1-hM4Di mice were not significantly different from D1-mCherry mice (Fig. S7).

Fig. 4. iSPN activation increases choice stochasticity



(**a**) Acquisition and Test phase trial history data from D2-Cre DREADD mice and controls were modeled using an RL model, and best fit parameters were inferred using hierarchical Bayesian model fitting. (**b**) Test phase alpha and β parameters did not significantly differ among

manipulation groups ($p>0.05$ Kruskal Wallis ANOVA). (**c**) Test phase β was significantly lower in D2-hM3Dq group compared to mCherry control (*$p<0.05$ Kruskal Wallis ANOVA).

Discussion

In the present study, we found that manipulating iSPN activity had surprising effects on learned choice behavior that were not accounted for by the popular 'select/suppress' heuristic. We showed that chemogenetically inhibiting dSPNs or activating iSPNs impaired suppression of nonrewarded choices, whereas inhibiting iSPNs did not affect choice behavior. These behavioral results were predicted by the OpAL network model, in which choice is determined by the relative balance of direct and indirect pathway activity (31). RL model fits to OpAL simulated data and mouse behavioral data showed that activating iSPNs reduced the inverse temperature (β) parameter, consistent with more exploratory choice. In the context of our deterministic task, this manifested as an increase in the number of nonrewarded choices. Our computational and empirical data demonstrate that the combined output of direct and indirect pathways, and not the independent function of either, is critical for adaptive choice behavior.

Previous studies provide clear evidence that optogenetic stimulation of DMS iSPNs can drive aversion and inhibit movement (32, 33, 45, 46), suggesting that choice suppression might be an extension of indirect pathway 'no go' function. However, recording data collected during decision-making has shown that dSPNs and iSPNs are simultaneously active when animals choose an action (23, 35, 36). The 'select/suppress' heuristic accounts for this coactivation by suggesting that dSPNs select specific actions while iSPNs simultaneously suppress alternate actions. However, it is problematic to infer *in vivo* function from optogenetic simulation effects that override endogenous activity patterns, especially if learning is stored in corticostriatal synaptic weights (6, 47, 48). Therefore, in the current study we chose to use chemogenetic manipulation tools in order to preserve aspects of endogenous activity that may have been altered during Acquisition phase learning.

While many studies have focused on the role of the striatum in selecting rewarded actions

(16, 18, 25) and stimuli (49), fewer have studied its role in avoiding low-value actions (50-53) and stimuli (54). Here, our goal was to understand how activity in DMS dSPNs and iSPNs influences choice behavior, particularly the ability to suppress an initially encountered low-value choice in order to make a subsequent high-value choice. In our odor-guided serial choice task, as in many natural decision-making settings, the value of a given action/choice (here, dig) was contingent on available stimulus information plus choice and outcome history. RL model fits to odor choice trial histories enabled us to investigate how chemogenetic manipulation altered the relationship between odor value estimates and choice. The multiple choice task design enhanced our ability to interpret underlying choice processes. Also, the fact that the mice were freely moving and were never required to hold still, thus removing the potential confound between choice suppression and motor freezing. Lastly, this task was acquired in a single session without extensive training that is often found in rodent operant tasks. This should enhance relevance to DMS function, which is engaged in early flexible goal-directed learning (22, 55, 56). Collectively, these more ethological task features may have permitted novel observations about the role of the indirect pathway in choice suppression behavior (57).

Our data support and add new circuit dimension to previously proposed dopaminergic mechanisms underlying choice exploration. We found that when iSPNs were activated (in chemogenetic manipulation data and OpAL simulations) choice became more stochastic/exploratory, meaning that mice were more likely to “explore” (i.e. choose) a lower value odor as opposed to “exploit” the highest value odor, as estimated by RL model fits. This was captured by a lower inverse temperature parameter, which tunes explore/exploit balance in the estimated odor value to choice conversion. This observation is consistent with a previous study that found D2R antagonism in the primate caudate reduced the inverse temperature parameter and increased exploratory choice (58). Our findings are also compatible with computational accounts that predict that lowering tonic dopamine, which facilitates iSPN activity and suppresses dSPN activity (59, 60), shifts explore/exploit balance towards exploration (61, 62), but see (63)

for an alternate model. Finally, if behavioral switching is viewed as exploring action space, our iSPN data may relate to recent studies that report iSPN activity increases in response to outcomes preceding switch trials (25, 64).

Unexpectedly, we observed that dSPN inhibition or iSPN activation increased the number of trials in which mice chose the first odor they encountered (Fig. 3). Several observations suggest that these single entry trials were separable from changes in locomotion and were not due to random impulsivity. Chemogenetic manipulation did not affect rotarod performance and the effects of manipulation on spontaneous locomotion in the open field did not correlate with entries in the serial choice task (Fig. 3). Regarding single entry trials, we reasoned that, if purely impulsive, the odors chosen on those trials would be random, i.e. independent of odor Q values. However, we found that single entry trial choices were significantly influenced by Test phase odor Q values (Figure S3). Therefore, we interpret the increase in single entry trials to be the result of a more exploratory choice process that occurs when chemogenetic inhibition of dSPNs or chemogenetic activation of iSPNs minimizes the difference in choice weights across odors.

It is possible that there are latent variable(s) in our task that are not captured by OpAL or our current RL models. For example, reduced entries prior to choice could reflect changes in cost/accuracy tradeoff and share mechanistic overlap with individuals with Parkinson's disease who are capable but choose not to exert the effort required to move rapidly in a motor speed/accuracy tradeoff task (65, 66), consistent with dissociable cognitive and motor impairments (67). Similarly, reduced entries may relate to the putative role of DMS in invigorating actions on the basis of net expected return and state value signals (68, 69). In addition, in our odor-guided serial choice task, reward contingency was 100%, and negative feedback was signaled by the absence of reward as opposed to punishment. OpAL predicts that inhibition of the indirect pathway more heavily influences choice behavior in environments in which animals balance reward and punishment or are rewarded in a probabilistic manner (31). Therefore, the deterministic nature of the task used here may have emphasized the contribution of the dSPNs

over iSPNs, potentially explaining why inhibiting iSPNs produced no detectable effects in this task.

Overall, our data support existing models of basal ganglia function in which trial and error choice drives learning that is later stored or read out in the balance of activity emerging from DMS dSPNs and iSPNs (70). The fact that learned choice behavior is specifically disrupted by chemogenetic inhibition of dSPNs and activation of iSPNs (but not by inhibition of iSPNs) is consistent with these manipulations counteracting reported patterns of long term potentiation (LTP) onto dSPNs and long-term depression (LTD) onto iSPNs following goal-directed action learning (6). Further work will need to be done to inform how LTP and LTD are allocated to specific neural ensembles of dSPNs and iSPNs to sculpt choice.

In summary, our findings suggest that the indirect pathway does not independently mediate choice suppression. Instead, choice appears to arise from the difference in dSPN and iSPN population activity, and conditions that reduce this difference increase choice stochasticity/exploration. Importantly, we demonstrate that manipulations that simply enhance activity in the indirect pathway do not facilitate adaptive choice suppression, and in fact can have the opposite effect. These data highlight the importance of using network concepts and models over simple heuristic accounts of circuit function to understand decision-making. We are hopeful that these findings will inform studies of addiction and other conditions in which enhancement of capacity for choice suppression is desirable.

Acknowledgments

We thank Yuting Zhang, Satya Vedula, Christopher Hall, and Nana Okada for assistance with behavior and histology. We thank Dr. Richard Ivry for manuscript feedback and Wilbrecht and Collins lab members for helpful discussion. This research was supported by a National Institute of Mental Health postdoctoral fellowship under Grant F32MH110184 to K.D.

Author contributions

K.D., A.G.E.C., and L.W. designed the research. K.D. performed all experiments and analyzed data. A.G.E.C. contributed analytic tools, designed and performed OpAL simulation and RL model analysis to which B.H. contributed. K.D., A.G.E.C., and L.W. wrote the manuscript.

Declaration of interests

The authors declare that there are not conflicts of interest.

Methods:

Mice

All mice were weaned on postnatal day (P)21 and group-housed on a 12:12hr reverse light:dark cycle (lights on at 10PM). C57BL/6 BAC transgenic mice expressing Cre recombinase under the regulatory elements for the D1 and D2 receptor (Drd1a-Cre and D2-Cre ER43) were obtained from Mutant Mouse Regional Resource and bred in our colony. Mice had ad lib access to food and water before food restriction in preparation for training. All procedures were approved by the Animal Care and Use Committee of the University of California, Berkeley and complied with the NIH guide for the use and care for laboratory animals.

Viruses and tracers

Adeno-associated viruses (AAVs) were produced by the Gene Therapy Center Vector Core at the University of North Carolina at Chapel Hill or by Addgene viral service and had titers of $>10^{12}$ genome copies per mL. For chemogenetic manipulations, mice were bilaterally injected with 0.5 uL of rAAV8-hsyn-DIO-mCherry, rAAV8-hsyn-DIO-hM3Dq-mCherry, or rAAV8-hsyn-DIO-hM4Di-mCherry. For *in vitro* electrophysiological validation experiments of rAAV8-hsyn-DIO-hM4Di-mCherry, mice were bilaterally injected with 0.69 uL of a 2:1 mixture of rAAV8-hsyn-DIO-hM4Di-mCherry and rAAV5-Ef1α-DIO-hChR2-EYFP.

Stereotaxic injections

Male and female mice (6-8 weeks) were deeply anesthetized with 5% isoflurane (vol/vol) in oxygen and placed into a stereotactic frame (Kopf Instruments; Tujunga, CA) upon a heating pad. Anesthesia was maintained at 1-2% isoflurane during surgery. An incision was made along the midline of the scalp and small burr holes were drilled over each injection site. Virus or tracer was delivered via microinjection using a Nanoject II injector (Drummond Scientific Company; Broomall, PA). Injection coordinates for DMS were (in mm from bregma): 0.90 anterior, +/-1.4 lateral, and -3.0 from surface of the brain. Injection coordinates for SNr were: 3.2 posterior, 1.2 lateral, and -4.6 from the surface of the brain. Mice were given subcutaneous injections of meloxicam (10 mg/kg) during surgery and 24 & 48 hours after surgery. Mice were group-housed before and after surgery and 4-6 weeks were allowed for viral expression before behavioral training or electrophysiology experiments.

Drugs

Clozapine-N-Oxide was generously provided by the NIMH Chemical Synthesis and Drug Supply Program (NIMH C-929). CNO was made fresh each day and dissolved in DMSO (0.5% final concentration) and diluted to 0.1 mg/mL in 0.9% saline USP. Tetrodotoxin (TTX), D-AP5, and NBQX disodium salt were purchased from Tocris Biosciences (Ellisville, MO).

Electrophysiology

Mice were deeply anesthetized with an overdose of ketamine/xylazine solution and perfused transcardially with ice-cold cutting solution containing (in mM): 110 choline-Cl, 2.5 KCl, 7 $MgCl_2$, 0.5 $CaCl_2$, 25 $NaHCO_3$, 11.6 Na-ascorbate, 3 Na-pyruvate, 1.25 NaH_2PO_4, and 25 D-glucose, and bubbled in 95% O_2/ 5%CO_2. 300 μm thick sections (sagittal for optogenetic stimulation experiment, coronal for all others) were cut in ice-cold cutting solution before being transferred to ACSF containing (in mM): 120 NaCl, 2.5 KCl, 1.3 $MgCl_2$, 2.5 $CaCl_2$, 26.2 $NaHCO_3$, 1 NaH_2PO_4

and 11 Glucose. Slices were bubbled with 95% O2/ 5% CO2 in a 37°C bath for 30 min, and allowed to recover for 30 min at room temperature before recording. All recordings were made using a Multiclamp 700B amplifier and were not corrected for liquid junction potential. The bath was heated to 32°C for all recordings. Data were digitized at 20 kHz and filtered at 1 or 3 kHz using a Digidata 1440 A system with pClamp 10.2 software (Molecular Devices, Sunnyvale, CA, USA). Only cells with access resistance of <25 MΩ were retained for analysis. Access resistance was not corrected. Cells were discarded if parameters changed more than 20%. Data were analyzed using pClamp or R (RStudio 0.99.879; R Foundation for Statistical Computing, Vienna, AT).

Spontaneous spiking in GPe neurons was recorded in cell-attached configuration. To evoke synaptic transmission by activating ChR2, we used a single wavelength LED system (470 nm; Thorlabs; Newtown, NJ) connected to the epifluorescence port of the Olympus BX51 microscope. Light pulses of 1-10 ms triggered by a TTL (transistor-transistor logic) signal from the Clampex software (Molecular Devices; Sunnyvale, CA) were delivered through a 63x objective and used to evoke synaptic transmission. Blue light pulses were delivered once every 10 s, and a minimum of 30 trials were collected. Light-evoked IPSCs were recorded in whole-cell configuration at +10 mV holding potential in the presence of D-AP5 (50 μM) and NBQX disodium salt (33 μM) to block glutamatergic neurotransmission. Recording pipettes had 2.5-5.5 MΩ resistances and were filled with internal solution (in mM): 115 Cs-methanesulfonate, 10 HEPES, 10 BAPTA, 10 Na2-phosphocreatine, 5 NaCl, 2 MgCl2, 4 Na-ATP, 0.3 Na-GTP.

Whole-cell current clamp recordings were performed using a potassium gluconate-based intracellular solution (in mM): 140 K Gluconate, 5 KCl, 10 HEPES, 0.2 EGTA, 2 MgCl2, 4 MgATP, 0.3 Na2GTP, and 10 Na2-Phosphocreatine. For current clamp recordings to validate CNO induced depolarization in Gq-DREADD- expressing Drd2+ neurons, ACSF contained 0.5 μM TTX and a stable baseline was collected for 3-5 minutes before ACSF containing 0.5 μM TTX + 10 μM

CNO was washed on. For all electrophysiology experiments, both male and female mice were used.

Behavioral assays

Adult male and female mice (6-10 weeks) were used in behavioral assays. Mice were first tested in 4 choice odor-guided serial choice task and then ≥ 2 weeks later were tested in locomotor and/or rotarod tasks so that performance on CNO could be compared within animals across tasks. Prior to all behavior assays, mice were habituated to the testing room for 30 minutes, and all behavior testing began 30 min after CNO treatment. Importantly, all groups (including DIO-mCherry) were administered CNO to control for potential off-target effect of the CNO metabolite clozapine (71).

4 choice odor-guided serial choice task:

The odor-guided serial choice task used has previously been described in detail (38, 72). In this task only the odor cue is predictive, and spatial or egocentric information are irrelevant. This behavior is also ethologically relevant because mice use odor information to locate food sources (73). Briefly, mice were food restricted to ~85 % bodyweight prior to training. On day 1, mice were habituated to the testing arena, on day 2 were taught to dig for cheerio reward in a pot filled with unscented wood shavings, on day 3 underwent a 4-choice odor discrimination in which they acquire the rule that 1 of 4 presented odors is rewarded, and finally on day 4 were tested for recall of the previously learned odor-reward association (Figure 2A). During the Test phase of the task, mice learned to discriminate among four pots with different scented wood shavings (anise, clove, litsea and thyme). All 4 pots were sham-baited with cheerio (under wire mesh at bottom) but only one pot was rewarded (anise). The pots of scented shavings were placed in each corner of an acrylic arena (12”, 12”, 9”) divided into 4 quadrants. Mice were placed in a cylinder in the center of the arena, and a trial started when the cylinder was lifted. Mice were then free to explore the

arena until a choice was signaled by a dig to the wood shavings. The cylinder was lowered as soon as a choice was made. If the choice was incorrect, the trial was terminated and the mouse was gently encouraged back into the start cylinder. Trials in which no choice was made within 3 minutes were considered omissions. If mice omitted for 2 consecutive trials, they received a reminder: a baited pot of unscented wood shavings was placed in the center cylinder and mice dug for the “free” reward. Mice were disqualified if they committed 4 pairs of omissions. The location of the 4 odors was shuffled on each trial, and criterion was met when the mouse completed 8 out of 10 consecutive trials correctly. 24-hours after completing the Test phase, mice underwent a recall Test of the initial odor-reward rule to criterion. For chemogenetic manipulation experiments, mice were injected with saline 30 minutes prior to discrimination training and injected with CNO (1.0 mg/kg) 30 minutes prior to testing in recall. During Acquisition and Test phase, experimenters (blind to group) manually scored entries into each quadrant, latency to dig, and odor choices. Importantly, in all behavioral assays mice expressing mCherry control virus were also administered CNO.

OpAL model

To simulate the effect of DREADD manipulation on the cortico-basal-ganglia network, we used the OpAL model, which is an approximation of a biologically realistic neural network model that includes direct and indirect pathways. Extended details of the OpAL model can be found in (31). Briefly, the model is an actor-critic-like RL algorithm which assumes two sets of weights are being tracked, D and I (corresponding to the direct pathway and indirect pathways, respectively; initialized at 1.5 for the preferred odor D weight, 1 for all other weights), in addition to a classic critic value V (initialized at 0.25 as the initial expected value of each option). Critic values are updated according to the classic RL equation:

$V_{t+1} \leftarrow V_t + alpha_C * delta_t$; $delta_t = r - V_t$ (r=1/0).

D and I weights are updated with a three factor, non-linear learning rule that emphasizes gains and losses, respectively:

- $D_{t+1} = D_t + [alpha_D * D_t] * delta_t$
- $I_{t+1} = I_t + [alpha_I * I_t] * [-delta_t]$
- To simplify the modeling of excitatory DREADD effects, we enforce an upper limit L to the D and I weights.
- The final choice is a softmax probability based on combined weights $Wt = beta_D * D_t - beta_I * I_t$, supposed to represent the output of the cortico-basal ganglia loops.

Chemogenetic tools modulate activity in dSPNs and iSPNs neurons, and we model their effects in the D and I weights respectively. We assume that inhibitory DREADD multiplicatively decrease the corresponding pathway's activity:

- $W_t = beta_D * DREADD_I * D_t - beta_I * I_t$ in the case of dSPN DREADD
- $W_t = beta_D * D_t - beta_I * DREADD_I * I_t$ in the case of iSPN DREADD

For the simulations, the inhibitory $DREADD_I$ parameter is fixed at 0.5.

For excitatory DREADD, we assume that the tool promotes firing in neurons whose activity would otherwise be subthreshold, thus increasing low weights more than high weights. Specifically, we model W activity as:

- $W_t = beta_D * D_t - beta_I * (I_t + DREADD_E * (L - I_t))$ in the case of iSPN DREADD

Where L is the activity limit. For all simulations, L is fixed at 2, and the excitatory $DREADD_E$ parameter is fixed at 0.8.

The choice between different odor options is given as a softmax choice policy on the linear combination of the choice weights. Therefore, when the choice weight for a candidate odor is much higher compared to the other odors, the policy is more exploitative. If choice weights across candidate odors are similar in value, the resulting policy is more exploratory. Because choice weights are the weighted difference between D and I weights, we can more simply state that choice policy is modulated by asymmetry between direct and indirect pathway activity.

To investigate the effects of chemogenetic manipulation on behavior, we simulated 100 times with parameters set to $alpha_C=alpha_D=alpha_I=0.1$, randomly chosen parameters $1<beta_D<3$, and $1<beta_I<1.6$, reflecting greater influence of the direct than indirect pathway on the final choice. We analyzed TTC as a function of DREADD condition, and show that it reproduces behavioral effects.

To interpret the model fit to the real data, we also fit the behavior of the BG model with the RL model fit to the mice (Figure 4A), using a similar – but non-hierarchical, standard fitting procedure (44). We find that the biologically realistic modeling of DREADD activity via the DREADD parameter predicts fit parameters with increased or decreased recall phase softmax β, as observed in the iSPN data.

Locomotor assay:

On day 1, mice underwent a habituation session in which they were placed in a clear acrylic box (225 x 225 mm) inside a sound attenuated chamber (Med Associates; Fairfax, VT) with lights off. Locomotion was monitored for 15 minutes using infrared beam breaks (Versamax, AccuScan Instruments, Columbus, OH). On days 2 and 3 mice received injections of saline or CNO (counterbalanced across mice) 30 minutes before their locomotion was monitored for 15 minutes. The chamber was cleaned with 70% ethanol between mice.

Rotarod test:

Females and males were run during separate sessions. On day 1, mice underwent a habituation trial in which they were placed individually in a clean holding cage for 5 mins. The rotarod (47650 Rota-Rod NG Ugo Basile; Monvalle VA, Italy) was then set at 5 rpm constant speed and each mouse was placed on the rod for 1 minute. The mice were then returned to the holding cage for another 5 mins before initiating the first trial. Each session consisted of 5 trials in which the rotarod constantly accelerated from 5-40 rpm over a period of 300 secs, and the latency at which mice

fell off the or held onto the rod for a full rotation was recorded. Mice rested for 5 mins in the holding cage between trials. Asymptotic performance was reached by day 3 of training (Supplementary Figure 5). On day 4, DIO-DREADD and DIO-mCherry mice were administered CNO (1 mg/kg, i.p.) 30 minutes before rotarod testing began. On day 5 mice were tested drug-free in rotarod performance. The rotarod apparatus was cleaned between mouse cohorts with 3% hydrogen peroxide (for plastic components) and 70% ethanol (for metal troughs).

Histology

Mice were transcardially perfused with PBS followed by 4% PFA in PBS. Following 24h postfixation, coronal brain slices (75 μm) were sectioned using a vibratome (VT100S Leica Biosystems; Buffalo Grove, IL). To confirm viral targeting, we performed a standard immunohistochemical procedure using a primary antibody against red fluorescence protein (RFP) (rabbit, Rockland 600-401-379; 1:1000) to enhance the mCherry signal expressed in mice transduced with rAAV8-hSyn-DIO-DREADD-mCherry or rAAV8-hSyn-DIO-mCherry. Sections were counterstained with DAPI (Life Technologies; Carlsbad, CA). Images were acquired with a Zeiss Axio Scan.Z1 epifluorescence microscope (Molecular Imaging Center, UC Berkeley) at 10x magnification and viewed using FIJI (ImageJ). For colocalization experiments, mCherry signal was enhanced as previously described, and images were acquired using a Zeiss LSM 710 confocal microscope (Biological Imaging Facility, UC Berkeley). Anatomical regions were identified according to the Mouse Brain in Stereotaxic Coordinates by Franklin and Paxinos and the Allen Institute Mouse Brain Atlas.

RL model

We modeled Acquisition and Test phase behavior using a reinforcement learning model driven by an iterative error-based rule (74, 75). The model uses a prediction error (δ) to update the value (V) of each odor stimulus. The prediction error is the difference between the experienced

feedback (λ) and the current expected value, where λ is 100 for rewarded choices and is 0 for incorrect choices. The prediction error is scaled by a learning rate parameter (α), with $0<\alpha<1$. Because mice exhibit innate preferences for odors, we set initial odor values to fixed parameters [v1,v2,v3,v4] for all 75 mice tested. Odors were selected so that the rewarded odor (O1, anise) was not the initially preferred odor. We confirmed that the choice distributions for each experimental group (mCherry, D2-hM4Di, D2-hM3Dq, and D1-hM4Di) did not significantly differ from the pooled average using a Chi-square test (data not shown).

To model trial-by-trial choice probabilities, the stimulus values were transformed using a softmax function to compute the relative probability of each choice. The inverse temperature parameter (β) determined the stochasticity of the choices. We used hierarchical Bayesian model fitting to infer the best fitting parameters, using the package STAN in Matlab (76). We assumed that odor values were shared by all animals, and that other parameters (α and β for each phase) were drawn from group level distributions defined by the experimental manipulation. We performed statistical tests on the distribution of samples obtained for the group-level hyperparameters. We compared the alternative models using the WAIC (77, 78), beginning with the simplest model (single α and β parameters shared across Acquisition and recall Test phases). We evaluated alternative models that included a phase decay parameter that allowed learned Q-values to decay to their initial values between the Acquisition and Test phases. In addition, we tested a model in which the Test phase α was set to =0, consistent with Q values not being updated during Test phase. We found that the best fit model included phase specific (non-zero) α and β parameters; all RL model comparisons are presented in Supplementary Table 1. We validated the models' parameter recovery and model comparison procedures on surrogate, simulated data. To validate the models and test how successfully each model captured the behavior of each mouse, we ran 100 task simulations for each mouse using the fit parameters. We then compared actual measures (e.g. trials to criterion) to the simulated average.

Statistics

Statistical tests were performed with GraphPad Prism 7.0 (San Diego, CA) and the R programming environment. For serial choice and locomotor behavioral data, groups were compared using one-way ANOVA if data were normally distributed or Kruskal Wallis test if data were not normally distributed. When the ANOVA or Kruskall Wallis test yielded significant results ($p < 0.05$), a post-hoc LSD or Dunn's test was used to compare DREADD manipulation groups to the mCherry control group. Because our experiments were designed to compare the behavior of DREADD manipulation groups to that of mCherry controls (planned comparisons), we did not correct for multiple comparisons.

References

1. Gillan CM*, et al.* (2015) Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *Am J Psychiatry* 172(3):284-293.
2. Kessler RM, Hutson PH, Herman BK, & Potenza MN (2016) The neurobiological basis of binge-eating disorder. *Neurosci Biobehav Rev* 63:223-238.
3. Lucantonio F, Stalnaker TA, Shaham Y, Niv Y, & Schoenbaum G (2012) The impact of orbitofrontal dysfunction on cocaine addiction. *Nature neuroscience* 15(3):358-366.
4. Yin HH, Ostlund SB, Knowlton BJ, & Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci* 22(2):513-523.
5. Yin HH, Knowlton BJ, & Balleine BW (2005) Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur J Neurosci* 22(2):505-512.
6. Shan Q, Ge M, Christie MJ, & Balleine BW (2014) The acquisition of goal-directed actions generates opposing plasticity in direct and indirect pathways in dorsomedial striatum. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 34(28):9196-9201.
7. Balleine BW & O'Doherty JP (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.
8. Balleine BW, Delgado MR, & Hikosaka O (2007) The role of the dorsal striatum in reward and decision-making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27(31):8161-8165.
9. Castane A, Theobald DE, & Robbins TW (2010) Selective lesions of the dorsomedial striatum impair serial spatial reversal learning in rats. *Behav Brain Res* 210(1):74-83.
10. Foerde K, Steinglass JE, Shohamy D, & Walsh BT (2015) Neural mechanisms supporting maladaptive food choices in anorexia nervosa. *Nature neuroscience* 18(11):1571-1573.

11. Volkow ND & Morales M (2015) The Brain on Drugs: From Reward to Addiction. *Cell* 162(4):712-725.
12. Everitt BJ*, et al.* (2008) Review. Neural mechanisms underlying the vulnerability to develop compulsive drug-seeking habits and addiction. *Philos Trans R Soc Lond B Biol Sci* 363(1507):3125-3135.
13. Friedman A*, et al.* (2017) Chronic Stress Alters Striosome-Circuit Dynamics, Leading to Aberrant Decision-Making. *Cell* 171(5):1191-1205 e1128.
14. Cox J & Witten IB (2019) Striatal circuits for reward learning and decision-making. *Nat Rev Neurosci*.
15. Hikosaka O, Nakamura K, & Nakahara H (2006) Basal ganglia orient eyes to reward. *J Neurophysiol* 95(2):567-584.
16. Lau B & Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* 58(3):451-463.
17. Kim H, Sul JH, Huh N, Lee D, & Jung MW (2009) Role of striatum in updating values of chosen actions. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29(47):14701-14712.
18. Tai LH, Lee AM, Benavidez N, Bonci A, & Wilbrecht L (2012) Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nature neuroscience* 15(9):1281-1289.
19. Donahue CH, Liu M, & Kreitzer AC (2018) Distinct value encoding in striatal direct and indirect pathways during adaptive learning. *bioRxiv*.
20. Gerfen CR*, et al.* (1990) D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* 250(4986):1429-1432.
21. Samejima K, Ueda Y, Doya K, & Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310(5752):1337-1340.
22. Thorn CA, Atallah H, Howe M, & Graybiel AM (2010) Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* 66(5):781-795.
23. Isomura Y*, et al.* (2013) Reward-modulated motor information in identified striatum neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33(25):10209-10220.
24. Shin JH, Kim D, & Jung MW (2018) Differential coding of reward and movement information in the dorsomedial striatal direct and indirect pathways. *Nature communications* 9(1):404.
25. Nonomura S*, et al.* (2018) Monitoring and Updating of Action Selection for Goal-Directed Behavior through the Striatal Direct and Indirect Pathways. *Neuron* 99(6):1302-1314 e1305.
26. Seo M, Lee E, & Averbeck BB (2012) Action selection and action value in frontal-striatal circuits. *Neuron* 74(5):947-960.
27. Alexander GE & Crutcher MD (1990) Functional Architecture of Basal Ganglia Circuits - Neural Substrates of Parallel Processing. *Trends in neurosciences* 13(7):266-271.
28. Albin RL, Young AB, & Penney JB (1989) The functional anatomy of basal ganglia disorders. *Trends in neurosciences* 12(10):366-375.
29. Mink JW (1996) The basal ganglia: Focused selection and inhibition of competing motor programs. *Prog Neurobiol* 50(4):381-425.
30. Frank MJ, Seeberger LC, & O'Reilly R C (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* 306(5703):1940-1943.
31. Collins AGE & Frank MJ (2014) Opponent Actor Learning (OpAL): Modeling Interactive Effects of Striatal Dopamine on Reinforcement Learning and Choice Incentive. *Psychol Rev* 121(3):337-366.

32. Kravitz AV*, et al.* (2010) Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* 466(7306):622-626.
33. Kravitz AV, Tye LD, & Kreitzer AC (2012) Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nature neuroscience* 15(6):816-818.
34. Yttri EA & Dudman JT (2016) Opponent and bidirectional control of movement velocity in the basal ganglia. *Nature* 533(7603):402-+.
35. Cui G*, et al.* (2013) Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature* 494(7436):238-242.
36. Jin X, Tecuapetla F, & Costa RM (2014) Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nature neuroscience* 17(3):423-430.
37. Hikosaka O, Takikawa Y, & Kawagoe R (2000) Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiological reviews* 80(3):953-978.
38. Johnson CM, Peckler H, Tai LH, & Wilbrecht L (2016) Rule learning enhances structural plasticity of long-range axons in frontal cortex. *Nature communications* 7:10785.
39. Parker JG*, et al.* (2018) Diametric neural ensemble dynamics in parkinsonian and dyskinetic states. *Nature* 557(7704):177-182.
40. Markowitz JE*, et al.* (2018) The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell* 174(1):44-58 e17.
41. Meng C*, et al.* (2018) Spectrally Resolved Fiber Photometry for Multi-component Analysis of Brain Circuits. *Neuron* 98(4):707-717 e704.
42. Barbera G*, et al.* (2016) Spatially Compact Neural Clusters in the Dorsal Striatum Encode Locomotion Relevant Information. *Neuron* 92(1):202-213.
43. Klaus A*, et al.* (2017) The Spatiotemporal Organization of the Striatum Encodes Action Space. *Neuron* 96(4):949.
44. Daw N (2009) Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning: Attention and Performance*, eds Delgado MR, Phelps EA, & Robbins TW (Oxford University Press), Vol XXIII, pp 1-23.
45. Freeze BS, Kravitz AV, Hammack N, Berke JD, & Kreitzer AC (2013) Control of basal ganglia output by direct and indirect pathway projection neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33(47):18531-18539.
46. Roseberry TK*, et al.* (2016) Cell-Type-Specific Control of Brainstem Locomotor Circuits by Basal Ganglia. *Cell* 164(3):526-537.
47. Xiong Q, Znamenskiy P, & Zador AM (2015) Selective corticostriatal plasticity during acquisition of an auditory discrimination task. *Nature* 521(7552):348-351.
48. Koralek AC, Jin X, Ii JDL, Costa RM, & Carmena JM (2012) Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. *Nature* 483(7389):331-335.
49. Santacruz SR, Rich EL, Wallis JD, & Carmena JM (2017) Caudate Microstimulation Increases Value of Specific Choices. *Curr Biol* 27(21):3375-3383 e3373.
50. Ogasawara T, Nejime M, Takada M, & Matsumoto M (2018) Primate Nigrostriatal Dopamine System Regulates Saccadic Response Inhibition. *Neuron* 100(6):1513-1526 e1514.
51. Bryden DW, Burton AC, Kashtelyan V, Barnett BR, & Roesch MR (2012) Response inhibition signals and miscoding of direction in dorsomedial striatum. *Front Integr Neurosci* 6:69.
52. Schmidt R, Leventhal DK, Mallet N, Chen F, & Berke JD (2013) Canceling actions involves a race between basal ganglia pathways. *Nature neuroscience* 16(8):1118-1124.
53. Friedman A*, et al.* (2015) A Corticostriatal Path Targeting Striosomes Controls Decision-Making under Conflict. *Cell* 161(6):1320-1333.
54. Kim HF, Amita H, & Hikosaka O (2017) Indirect Pathway of Caudal Basal Ganglia for Rejection of Valueless Visual Objects. *Neuron* 94(4):920-930 e923.

55. Kupferschmidt DA, Juczewski K, Cui G, Johnson KA, & Lovinger DM (2017) Parallel, but Dissociable, Processing in Discrete Corticostriatal Inputs Encodes Skill Learning. *Neuron* 96(2):476-489 e475.
56. Yin HH*, et al.* (2009) Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nature neuroscience* 12(3):333-341.
57. Juavinett AL, Erlich JC, & Churchland AK (2018) Decision-making behaviors: weighing ethology, complexity, and sensorimotor compatibility. *Current Opinion in Neurobiology* 49:42-50.
58. Lee E, Seo M, Dal Monte O, & Averbeck BB (2015) Injection of a dopamine type 2 receptor antagonist into the dorsal striatum disrupts choices driven by previous outcomes, but not perceptual inference. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 35(16):6298-6306.
59. Surmeier DJ, Ding J, Day M, Wang ZF, & Shen WX (2007) D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends in neurosciences* 30(5):228-235.
60. Mallet N, Ballion B, Le Moine C, & Gonon F (2006) Cortical inputs and GABA interneurons imbalance projection neurons in the striatum of parkinsonian rats. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26(14):3875-3884.
61. Humphries MD, Khamassi M, & Gurney K (2012) Dopaminergic Control of the Exploration-Exploitation Trade-Off via the Basal Ganglia. *Frontiers in neuroscience* 6:9.
62. Dunovan K & Verstynen T (2016) Believer-Skeptic Meets Actor-Critic: Rethinking the Role of Basal Ganglia Pathways during Decision-Making and Reinforcement Learning. *Frontiers in neuroscience* 10:106.
63. Sridharan D, Prashanth PS, & Chakravarthy VS (2006) The role of the basal ganglia in exploration in a neural model based on reinforcement learning. *Int J Neural Syst* 16(2):111-124.
64. Geddes CE, Li H, & Jin X (2018) Optogenetic Editing Reveals the Hierarchical Organization of Learned Action Sequences. *Cell* 174(1):32-43 e15.
65. Mazzoni P, Hristova A, & Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27(27):7105-7116.
66. Baraduc P, Thobois S, Gan J, Broussolle E, & Desmurget M (2013) A Common Optimization Principle for Motor Execution in Healthy Subjects and Parkinsonian Patients. *Journal of Neuroscience* 33(2):665-677.
67. Middleton FA & Strick PL (2000) Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain Cogn* 42(2):183-200.
68. Hamid AA*, et al.* (2016) Mesolimbic dopamine signals the value of work. *Nature neuroscience* 19(1):117-126.
69. Wang AY, Miura K, & Uchida N (2013) The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nature neuroscience* 16(5):639-+.
70. Bariselli S, Fobbs WC, Creed MC, & Kravitz AV (2018) A competitive model for striatal action selection. *Brain Res*.
71. Mahler SV & Aston-Jones G (2018) CNO Evil? Considerations for the Use of DREADDs in Behavioral Neuroscience. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*.
72. Johnson C & Wilbrecht L (2011) Juvenile mice show greater flexibility in multiple choice reversal learning than adults. *Dev Cogn Neurosci* 1(4):540-551.
73. Howard WE, Marsh RE, & Cole RE (1968) Food detection by deer mice using olfactory rather than visual cues. *Anim Behav* 16(1):13-17.

74. Sutton RS & Barto AG (1998) *Reinforcement learning : an introduction* (MIT Press, Cambridge, Mass.) pp xviii, 322 p.
75. Rescorla RA & Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, eds Black A & Prokasy W (Appleton Century Crofts., New York), pp 64-99.
76. Carpenter B*, et al.* (2017) Stan: A Probabilistic Programming Language. *J Stat Softw* 76(1):1-29.
77. Watanabe S (2010) Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J Mach Learn Res* 11:3571-3594.
78. Vehtari A, Gelman A, & Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 27(5):1413-1432.