# **Supplementary Figure 1**



(**a**) Additional measures of the reversal learning task

(left panel) Plot of the cumulative reversals over time for all animals after systemic drug injection, confirming that the drug-induced performance impairment does not develop until after the first reversal (dashed line). Sidak's multiple comparisons test: p < 0.05 after trial 58 for Damphetamine, p < 0.05 after trial 62 for cocaine. (right panels)

Trials rewarded: one-way repeated measures ANOVA,  $F(1.670, 40.08) = 3.998$ ,  $p = 0.0327$ . Post-hoc Sidak's test: cocaine vs saline,  $t(24) = 2.358$ ,  $p = 0.0530$ ; D-amphetamine vs saline,  $t(25) = 1.561$ ,  $p = 0.2461$ .

Time to complete session: one-way repeated measures ANOVA,  $F(1.930, 46.33) = 3.454$ ,  $p =$ 0.0415. Post-hoc Sidak's test: cocaine vs saline,  $t(24) = 2.388$ ,  $p = 0.0497$ ; D-amphetamine vs saline,  $t(25) = 0.2042$ ,  $p = 0.9744$ .

Data shows mean ± standard error of the mean.

(**b**) Likelihood of model fits. Every dot represents an individual session. Cocaine and Damphetamine did not significantly affect the fit of the model to the data (one-way repeated measures ANOVA, F(2, 48) = 1.783, p = 0.1791).

(**c**) Heatplot of simulated data showing how win- and lose-stay behavior (taken over the entire session) vary as a function of learning rates  $a_{win}$  and  $a_{loss}$ . Data shown are the average of 100 simulations of each  $q_{\text{win}}/q_{\text{loss}}$  combination, with choice stochasticity factor  $\beta$  fixed at its mean for visualization purposes ( $β = 6.7$ ). Dashed black lines show the average estimated learning rates after saline injection. The win-stay parameter is relatively stable for high learning rates compared to lose-stay, while lose-stay is more stable for lower learning rates. Hence, a decline of the average negative learning rate  $q_{loss}$  by  $\sim$ 2/3 more strongly affects win-stay than lose-stay behavior, providing an explanation for the observation that cocaine and D-amphetamine affect win-stay, but not lose-stay behavior. In contrast, when baseline learning rates would have been high, a decrease in  $q_{\text{loss}}$  would have resulted in an increase in lose-stay, without affecting winstay behavior. Thus, how learning rates affect win- and lose-stay behavior is dynamic, and this strongly depends on the baseline estimates of α<sub>win</sub>, α<sub>loss</sub> and β.



(**a**) (left) Spread of expression of Gq-mCherry in the midbrain. Shown is -5.40 mm posterior to Bregma. Atlas image adapted from Supplementary reference 1.

(right) Quantification of number of Gq-mCherry transfected neurons per group. Each dot represents a single animal. Significantly fewer neurons were transfected in the mesocortical group compared to the mesoaccumbens group (unpaired t-test,  $t(14) = 6.713$ ,  $p < 0.0001$ ). (**b**) Quantification of expression of Gq-mCherry in the midbrain. In mesoaccumbens animals, virus sometimes spread to the medialmost part of the substantia nigra (SN), although this was always less than 5% of total transfected neurons.

(**c**) Example histology image of an animal from the mesoaccumbens group, showing strong expression of Gq-mCherry in the VTA and modest expression in the medial SN.



(a) No effect of CNO treatment on the cumulative reversals over time for the control group and the mesocortical group (two-way repeated measures ANOVA for control group: main effect of CNO, F(1, 16) = 2.919, p = 0.1068; trials × CNO interaction, F(149, 2384) = 0.7633, p = 0.9838; two-way repeated measures ANOVA for data mesocortical group: main effect of CNO, F(1, 15) = 0.2858, p = 0.6007; trials  $\times$  CNO interaction, F(148, 2220) = 0.5058, p > 0.9999). The complete section  $\mathbb{N}$ <br>  $\mathbb{N}$  ;

(b) Lose-stay behavior during reversal learning is not affected by DREADD stimulation of either pathway.

Left: two-way repeated measures ANOVA; main effect of CNO,  $F(1, 40) = 0.1325$ ,  $p = 0.7178$ ; group  $\times$  CNO interaction, F(2, 40) = 2.136, p = 0.1314.

Right: two-way repeated measures ANOVA; main effect of CNO,  $F(1, 50) = 1.392$ ,  $p =$ 0.2436; group  $\times$  CNO interaction, F(2, 50) = 0.045, p = 0.9556.

(**c**) (left panel) Model fit on the reversal learning data of the mesoaccumbens group. DREADD activation altered  $q_{\text{loss}}$  in the same direction as cocaine and D-amphetamine, although not significantly so (one-tailed Wilcoxin matched-pairs signed rank test, W = -41.00,  $p = 0.1764$ ).

(right panel) Mesoaccumbens activation resulted in a significantly poorer fit of the model to the data (paired t-test,  $t(16) = 3.224$ ,  $p = 0.0053$ ). This seems consistent with the observation that during mesoaccumbens hyperactivity, both win-stay (Fig. 2e) and lose-stay behavior (Supplementary Figure 3b) are around chance level (50%), making the Rescorla-Wagner model a suboptimal descriptor of the animals' behavior.



*In vivo* microdialysis performed in the NAc showed increased baseline levels of DA and its metabolites after activation of the mesoaccumbens pathway by CNO (n = 4 animals DREADD  $aroup$ ,  $n = 5$  animals control group)

Two-way repeated measures ANOVA, with factors treatment and timepoints: DA:

Main effect of treatment:  $F(1,7) = 11.83$ ,  $p = 0.0108$ 

Treatment  $\times$  Time interaction effect:  $F(9,63) = 4.11$ ,  $p = 0.0003$ 

DOPAC:

Main effect of treatment:  $F(1,7) = 9.77$ ,  $p = 0.0167$ 

Treatment  $\times$  Time interaction effect:  $F(9,63) = 15.69$ ,  $p < 0.0001$ 

HVA:

Main effect of treatment:  $F(1,7) = 9.01$ ,  $p = 0.0199$ 

Treatment  $\times$  Time interaction effect:  $F(9,63) = 23.65$ ,  $p < 0.0001$ 

Post-hoc LSD tests: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ ). Note a possible type I error at time point 3 in the DA graph.



b



Animals photometry (figure 3;  $n = 5$ )<br>  $\circ$  Animals photometry-DREADD combination (figure 7;  $n = 4$ )

Fiber photometry

(**a**) Photometry responses during reversal learning in animals injected with the control fluorophore AAV-hSyn-eYFP (mean ± standard error of the mean).

(**b**) Fiber placement of animals used in photometry recordings. Atlas image adapted from Supplementary reference 1.



b







Probabilistic discounting task.

(**a**) In the probabilistic discounting task with decreasing probabilities across trial blocks, responding on the risky lever is economically beneficial in the first block, responding on the safe lever is beneficial in the last two blocks. In the second block, the yield of both levers is equal. The opposite is true for the version of the task with increasing probabilities across trial blocks.

(**b**) Depending on a priori knowledge, in the first block of the probabilistic discounting task, deand revaluative mechanisms are needed to determine the reward value of the safe and risky levers. Assuming that a proper neuronal representation of lever value has been established at the end of the first block, subsequent blocks in the probabilistic discounting task with decreasing probabilities (left column) involve devaluative mechanisms, whereas the probabilistic discounting task with increasing probabilities (right column) involve revaluative mechanisms.

(**c**) Safe-stay behavior, defined as the percentage of safe choice trials followed by another safe choice, was unaffected by CNO treatment (two-way repeated measures ANOVA, main effect of CNO, F(1, 34) = 1.050, p = 0.3127; group  $\times$  CNO interaction, F(2, 34) = 1.365, p = 0.2690).

(**d**) Percentage choice of the risky lever in the probabilistic discounting task with increasing probabilities during mesoaccumbens stimulation. Only in the first 5 trials of block 1, mesoaccumbens activation increased the choice for the risky lever, despite the low chance on reward (Fisher's LSD test in block 1:  $t = 2.652$ ,  $p = 0.0096$ . In all other blocks:  $p > 0.2$ ).



#### Elevated plus maze

(**a**) Example track of a control animal in the elevated plus maze. Red line indicates the track of the animal's center point. Scalebar, 25 cm.

(**b**) Total time spent in the closed arms of the elevated plus maze. Stimulation of the mesocortical pathway showed a trend towards increased anxiety, whereas stimulation of the mesoaccumbens pathway had no effect on behavior (unpaired t-test with Welch's correction for unequal variance, Bonferonni corrected for 2 comparisons;  $F(26.22)$ uncorrected = 1.943, p = 0.1256 for mesoaccumbens versus control,  $F(21.25)$ uncorrected = 2.378,  $#p = 0.053$  for mesocortical versus control).  $n = 16$  control,  $n = 15$  mesoaccumbens,  $n = 17$  mesocortical.



<sup>(</sup>*Figure legends on next page)* 

(**a**) Mesoaccumbens stimulation increases locomotion (Sidak's multiple comparisons test, mesoaccumbens versus control,  $t(44) = 4.383$ ,  $p = 0.0001$ ; mesocortical versus control,  $t(44)$  $= 0.1096$ ,  $p = 0.9925$ ). All animals received CNO.

(**b**) Reaction times in the punishment task (based on trials 11-30). Receiving a foot shock during a trial robustly increased the reaction time during the subsequent trial in all three groups (two-way repeated measures ANOVA; main effect of shock, all groups p < 0.01). In addition, a significant main effect of CNO ( $F(1,9) = 20.97$ ,  $p = 0.0013$ ) and a significant shock  $\times$  CNO interaction (F(1,9) = 8.271, p = 0.0183) were observed in the mesoaccumbens group. Post-hoc Sidak's multiple comparisons test revealed a significant slowing of responding after mesoaccumbens activation after a no-shock trial (t(9) = 4.532,  $p = 0.0028$ ), as well as after a shock trial  $(t(9) = 8.599, p < 0.0001)$ .

(**c**) Mesocortical or mesoaccumbens activation did not affect inactive nose poking in the punishment task (2-way repeated measures ANOVA; main effect of CNO, F(1,25) < 0.0001, p  $= 0.9946$ ; group  $\times$  CNO interaction, F(2, 25) = 0.3164, p = 0.7316).

(**d**) Time animals needed to complete the 150 trials of the reversal learning session was unaffected by CNO treatment (two-way repeated measures ANOVA; main effect of CNO, F(1,  $47$ ) = 0.0439, p = 0.8350; group  $\times$  CNO interaction, F(2, 47) = 0.2961, p = 0.7451).



Photometry recordings of VTA DA neurons during DREADD activation.

(**a**) Data from individual animals from figure 7a.

(b) To correct for bleaching, raw calcium signal was converted to dF/F<sub>0</sub> values by normalizing to a running-average baseline.

# **Supplementary tables**



**Supplementary table 1** Model fits, performed on baseline behavior (i.e., after saline treatment) in the reversal learning task in the  $n = 25$  rats from figure 1. Model 1 (' $M_1$ ') is the classical Rescorla-Wagner model, whereas model 2 ('M<sub>2</sub>') uses separate learning rates for reward ( $a_{win}$ ) and punishment ( $\alpha_{loss}$ ) learning. Since the tested models are nested ( $M_1$  is a special case of  $M_2$ ), model comparison was performed using the likelihood-ratio test.  $M_0$  is the baseline model, in which choice behavior is random ( $p = 0.5$  for every trial).



**Supplementary table 2** Best-fit model parameters, estimated by maximizing the log likelihood for the model given the choice sequences in every session. Wilcoxon matched-pairs signed rank test with Bonferroni correction,  $q_{loss}$ : cocaine versus saline,  $p = 0.0046$ ; D-amphetamine versus saline, *p* = 0.032. For additional statistics see the Methods checklist.

# **Supplementary table 3** Supplementary statistics















## **Supplementary references**

1. G. Paxinos and C. Watson. The Rat Brain in Stereotactic Coordinates (6th Edition). Elsevier Inc. (2007).