

Supplementary Methods:

Methods and Materials:

Animals: Male and female 6- to 14-week-old C57BL/6J (n = 75; Bar Harbor, ME; SN: 000664) or D1-Cre (n = 13; Jackson Laboratories; Bar Harbor, ME; #030329) mice were housed five per cage. All animals were maintained on a reverse 12h dark-12h light cycle. Animals had free access to water but were food restricted to 90% of free-feeding weight for the duration of the studies. Mice were weighed every other day to ensure that weight was maintained. Animals were fed 2.5g chow per/mouse/day and food intake was adjusted to meet the weight criteria based on animals' body weight each day. Behavior was conducted during the dark phase of the light cycle. All experiments were conducted in accordance with the guidelines of the Institutional Animal Care and Use Committee at Vanderbilt University School of Medicine, which approved and supervised all animal protocols. Experimenters were blind to experimental groups and the order of testing was counterbalanced during behavioral experiments.

Apparatus: Mice were trained and tested daily in individual Med Associates (St. Albans, Vermont) operant conditioning chambers fitted with a house light, grid floor with shock harness, programmable tone generator, speakers, and two illuminated nose-pokes on either side of a sucrose delivery port equipped with an infrared beam break to assess head entries. One nose-poke functioned as the active operanda, and the other as the inactive, depending on the phase of the experiment (described below). Responses on both nose-pokes and head entries into the sucrose port were recorded throughout the duration of the experiments.

Multidimensional cue outcome action task (MCOAT): *Experimental Timeline:* Animals were trained in a series of operant tasks, for which they must meet task-specific criteria (defined below) before moving to the next phase. Mice that fail to meet each criterion do not continue to the next phase - the percentage of animals in each group completing criteria is outlined in **Figure 2** and **Supplementary Fig 3**. The different phases of the task are (**Phase 1**) positive reinforcement and negative reinforcement, (**Phase 2a**) limited discrimination and conflict, (**Phase 2b**) extensive discrimination and conflict, (**Phase 3**) and punished responding (**Fig 1**). Animals were trained in one

1h session daily. All experiments were done within subjects allowing for comparisons to be made across training sessions and conditions.

Phase 1: Positive and Negative Reinforcement

Positive Reinforcement. Mice were trained on a fixed-ratio 1 (FR1) schedule of reinforcement to nose-poke in the active poke for sucrose delivery (1s duration of delivery, 10uL volume, 1mg sucrose; Fisher Scientific). Upon each correct response the sucrose delivery port was illuminated for 5 seconds and sucrose was delivered. During *Phase 1* training sessions, an auditory discriminative stimulus (**S_{d1}**) - white noise or 2.5 khz tone (counterbalanced) - was presented for the entirety of the session. Mice were moved to the next phase when they responded on the active NP >80 times in a session. Each session was 1 hr in total duration.

Negative Reinforcement. Mice were trained to nose-poke on the opposite, non-sucrose-paired nose-poke for negative reinforcement - to prevent the presentation of foot shocks. The order of positive and negative reinforcement was counterbalanced and there were no significant order effects. All shocks were short, but high intensity: 1.0 mA in magnitude delivered for 0.5s. A second auditory discriminative stimulus (**S_{d2}**) -- either tone or white noise, counterbalanced between positive and negative reinforcement -- was presented on a variable interval 30s (VI30) schedule for the inter-trial interval (ITI). At the beginning of each trial, **S_{d2}** came on for 30s after which a series of shocks was delivered (15 second inter-stimulus interval (ISI), 20 shocks total). In this task, mice are able to respond any time during the trial to end the trial and begin the ITI. Responding on the correct nose-poke during **S_{d2}** immediately ended the trial, thus preventing the shocks from being presented (avoidance). If responses were made after shocks commenced, responding on the correct nose-poke terminated the shocks and ended the trial (escape). The shock and **S_{d2}** were terminated immediately following a correct response. The trial ended either after the animal made a correct response or after 330 seconds. Each session was 1 hr in total duration. Unlike the positive reinforcement phase, we used discrete cues with a variable ITI for the negative reinforcement phase as the cue signals the presence of a potential outcome to be removed instead of an outcome to be obtained. Therefore, for successful learning negative reinforcement requires a time out period following a correct response, which is not necessary for positive reinforcement. Acquisition criteria was defined

as receiving fewer than 25% of total possible shocks in a session. Animals that did not meet this criterion after 15 sessions were removed from the study.

Phase 2a: Limited Discrimination and Conflict. Following the acquisition of both the positive and negative reinforcement tasks, mice went into *Phase 2a*. In the limited discrimination phase, mice were trained in one 1h session per day for three consecutive days. In this trial-based phase, 80% of the trials were discrimination trials and 20% were conflict trials.

Limited Discrimination Pre-Training: Before the beginning of the conflict testing, animals underwent three sessions of discrimination only training to ensure that they were using the antecedent cues (\mathbf{S}_{d1} OR 2) to guide their operant responses. In these trials, \mathbf{S}_{d1} and \mathbf{S}_{d2} were presented in random order and equal proportion and responses on the correct (corresponding to the \mathbf{S}_d that was presented) and incorrect nose-poke were recorded. The \mathbf{S}_d predicted the same response between phase 1 and 2, the only difference is that they were presented randomly within the same session to ensure that the animals had acquired the association. Depending on the \mathbf{S}_d , animals could respond on the appropriate nose-poke for either sucrose or shock avoidance. Response on the active nose-poke during \mathbf{S}_{d1} initiated a 1s sucrose delivery with 5s sucrose port illumination and terminated the \mathbf{S}_{d1} , effectively ending the trial. Response at the opposing nose-poke during \mathbf{S}_{d2} terminated \mathbf{S}_{d2} and ended the trial. Failure to make an active response during the 30 second duration of the \mathbf{S}_{d2} resulted in a single shock and the trial ended. Mice that did not respond in either sucrose or shock trials were removed during this phase.

Discrimination and Conflict: The test session consisted of both discrimination trials (80% of trials) and conflict trials (20% of trials) in the same session. Discrimination trials were identical to those described above. In conflict trials, mice were presented with a compound cue ($\mathbf{S}_{d1} + \mathbf{S}_{d2}$) for 30 seconds. Both nose-pokes were illuminated. Depending on their response, mice received one of three possible outcomes: 1) failure to respond resulted in a shock at the end of the 30s compound cue, 2) if they responded on the sucrose active side, they received sucrose and a footshock, 3) If they responded on the negative reinforcement active side, they avoided shock and did not receive sucrose. As before, trials and \mathbf{S}_d s were terminated following an active response. This allowed us to define animals' response bias when conflicting information was presented.

Phase 2b: Extensive Discrimination and Conflict. Following acquisition of both the positive and negative reinforcement tasks, a second cohort of mice underwent extensive discrimination training. Each day, mice underwent a 15 minute pre-discrimination positive reinforcement session, a 15 min pre-discrimination negative reinforcement session (0.3 mA, 0.5 sec shock), and a 1hr discrimination/conflict session (80% discrimination trials, 20% conflict trials as described above). The animals received the shock pre-discrimination trials first and sucrose trials second to make sure mice still seek for reward after getting shocked. In both the positive and negative reinforcement sessions, the mice responded in >80% of the trials to move onto the next session for that day. Mice were trained daily in discrimination until they reached a criterion of >70% correct.

Phase 3: Punished Responding. Mice trained in positive and negative reinforcement and that underwent limited discrimination and conflict (*Phase 2*) were moved to punished responding. Each session contained 50% positive reinforcement trials for sucrose and 50% punished trials.

Positive reinforcement trials: Mice were presented **S_{d1}** and had 30 seconds to nose-poke on the active poke for sucrose. Sucrose was delivered as described above. **S_{d1}** and the trial were terminated following an active response or at the end of the 30 seconds.

Punished trials: **S_{d1}** and **S_{d3}** (a house light) were presented concurrently. Responding on the active nose-poke on these trials resulted in the delivery of sucrose and a single footshock. The intensity of this shock was increased in each subsequent session over the course of 9 total sessions (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.75, 1.0, and 1.5 mA). In these trials to avoid shock the animal must inhibit behavioral responding - thus the outcome is the same as in phase 1 (avoid shock) but the behavior is opposite (go vs no go).

Shock Sensitivity: To rule out differences in shock sensitivity between males and females as a factor contributing to the behavioral outcomes, a follow-up shock sensitivity task was conducted after the punished responding phase. During one 1h session, animals received randomly selected magnitude shocks of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.75, 1.0, or 1.5 mA with variable ITI of 30, 45, or 60 seconds. All shocks were unsignaled and no cues were presented in the session. Vocalization (non-ultrasonic) and motor responses were scored. Vocalization was scored as a 1 if the subject vocalized and a 0 if the subject did not vocalize in the session.

Motor responses were scored as a 1 if the subject ran, a 2 if the subject hopped (4 paws off the ground), a 3 if the subject ran and hopped, and as a 0 if the subject did not move.

Chemogenetic inhibition experiments: For the D1 MSN chemogenetic inhibition experiments, we injected inhibitory designer receptors exclusively activated by designer drugs (DREADDs) into the nucleus accumbens (NAc). For these surgeries, Ketoprofen (5mg/kg; subcutaneous injection) was administered at least 1 hr before surgery. Under Isoflurane anesthesia, mice were positioned in a stereotaxic frame (Kopf Instruments) and the NAc was targeted (bregma coordinates: anterior/posterior, + 1.4 mm; medial/lateral, + 1.5 mm; dorsal/ventral, -4.3 mm; 10° angle). Using aseptic technique, a midline incision was made down the scalp and a craniotomy was made using a dental drill. A 10-mL Nanofil Hamilton syringe (WPI) with a 34-gauge beveled metal needle was used to infuse AAV2/hSyn-DIO(Gi)-hM4Di-mCherry (Addgene #44362). Virus was infused at a rate of 50 nL/min for a total of 500 nL. Following infusion, the needle was kept at the injection site for seven minutes and then slowly withdrawn. Follow up care was performed according to IACUC/OAWA and DAC standard protocol. Animals were allowed to recover for a minimum of four weeks to ensure efficient viral expression before commencing experiments.

For the behavioral inhibition experiments, we injected Clozapine N-oxide (CNO; 5 mg/kg) or saline 30 mins prior to the behavioral testing. Mice were tested in positive reinforcement, negative reinforcement, limited discrimination, and conflict phases of the MCOAT paradigm. CNO or saline was injected IP 30 mins prior to behavioral testing to inhibit D1 MSNs during each discrete phase. An experimental timeline denoting the testing sessions where CNO/saline was administered is presented in **Fig 6B**. Mice were injected with CNO (or saline as a control) before the first and the second trial of positive and negative reinforcement to determine how affected acquisition. Then once mice had acquired CNO/saline was administered during the last two sessions to determine the effects of D1 MSN inhibition on ongoing performance. All mice received a total of 14 positive and negative reinforcement sessions. Each mouse received a CNO and a saline injection and the order of the injections were counterbalanced (CNO-Saline or Saline-CNO). Therefore, we ensured that all mice received the same number of CNO injections albeit one group at the baseline conditions (Session 1) and another group during the initial learning phase (Session 2).

During the discrimination phase, mice were first given 2 drug-free discrimination sessions and received CNO or saline injections during the next 2 sessions. Finally, all mice received 2 conflict trials where they also received the CNO and saline injections in a counterbalanced manner.

Histology: Subjects were deeply anaesthetized with an intraperitoneal injection of Ketamine/Xylazine (100mg/kg/10mg/kg) and transcardially perfused with 10 mL of PBS solution followed by 10 mL of cold 4% PFA in 1x PBS. Animals were decapitated, the brain was extracted and placed in 4% PFA solution and stored at 4 °C for at least 48- hours. Brains were then transferred to a 30% sucrose solution in 1x PBS and allowed to sit until brains sank to the bottom of the conical tube at 4 °C. After sinking, brains were sectioned at 35µm on a freezing sliding microtome (Leica SM2010R). Sections were stored in a cryoprotectant solution (7.5% sucrose + 15% ethylene glycol in 0.1 M PB) at -20 °C until immunohistochemical processing. Sections were mounted on glass microscope slides with Prolong Gold antifade reagent. Fluorescent images were taken using a Keyence BZ-X700 inverted fluorescence microscope (Keyence), under a dry 20x objective (Nikon). The injection site location was determined with serial images in all animals.

Analysis Parameters: For positive and negative reinforcement tasks, the total sucrose and total shock responses were analyzed using unpaired t tests. Mann-Whitney U test was used when the number of sessions to the criterion was not equal between subjects precluding the use of parametric statistics. Discrimination and conflict task responses were analyzed using two-way ANOVA (Trial Type x Sex) when they were represented as averages or when they represent a single session (Limited Discrimination Phase). We employed a mixed Repeated Measures ANOVA for the Punished Responding and Shock Sensitivity experiments where all mice received the same number of sessions. We also used a computational analysis to determine the parameters of response bias (Log b) and discrimination (Log d), as described previously (1, 2). Briefly, Log d value was derived mathematically as a measure of the rate of discrimination in a bias-independent measure whereas Log b was computed as the measure for behavioral bias. Both terms use a logarithmic scale for the multiplication of the ratio between correct and incorrect responses during two different trial types. We explain the mathematical terms in detail below:

Log d: *Log d* is a measure of the rate of discrimination for the S_d . In the discrimination phase of the MCOAT, mice were trained to nose-poke in the right or left poke based on the S_d presented to them in order to obtain sucrose or avoid footshocks. *Log d* is determined as the ratio between the number of correct and incorrect Sucrose and Shock trials, which results in a negative (no discrimination) or a positive (successful discrimination):

$$\mathbf{Log\ d = 0.5 * log [((Sucrose_{correct}+0.5) * (Shock_{correct}+0.5)) / ((Sucrose_{incorrect}+0.5) * (Shock_{incorrect}+0.5))]}$$

Log b: *Log b* is a measure of the animals' response bias. In our task, the animals were presented a compound stimulus consisting of two auditory cues signaling opposite outcomes. In these conflict trials, the subjects had to choose between getting a sucrose reward versus avoiding a footshock. Using the data from these trials, we assessed "behavioral bias," that is, the preference of mice for one outcome over another. *Log b* is calculated as the ratio between the number of correct Sucrose and incorrect Shock versus incorrect Sucrose and correct Shock trials, which results in either a negative (bias towards avoidance) or a positive value (bias towards sucrose) or a 0 (no bias):

$$\mathbf{Log\ b = 0.5 * log [((Sucrose_{correct}+0.5) * (Shock_{incorrect}+0.5)) / ((Sucrose_{incorrect}+0.5) * (Shock_{correct}+0.5))]}$$

Additional References

1. Branch MN (1977): On the role of "memory" in the analysis of behavior. *J Exp Anal Behav.* . doi: 10.1901/jeab.1977.28-171.
2. Kangas BD, Berry MS, Branch MN (2011): On the Development and Mechanics of Delayed Matching-to-Sample Performance. *J Exp Anal Behav.* . doi: 10.1901/jeab.2011.95-221.