Title: Distinct roles for DAT and COMT in regulating dopamine transients and reward-guided decision making

Abbreviated title: Roles of DAT and COMT in dopamine and reward

Authors & affiliations: Clio Korn[1], Thomas Akam[2], Kristian HR Jensen[1], Cristiana Vagnoni[2], Anna Huber[1], Elizabeth M Tunbridge[1]**, Mark E Walton[2,3]**

** equal contribution

1. Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, UK

2. Department of Experimental Psychology, University of Oxford, OX2 6GG, Oxford, UK

3. Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, OX3 9DU, UK

corresponding authors: clio.korn@ucsf.edu; mark.walton@psy.ox.ac.uk

Page and word counts

Number of pages: 51

Number of figures: 5 (plus 4 extended figures)

Number of tables: 4

Number of words: Abstract 234, Introduction 678, Discussion 1557

Conflict of interest statement: The authors declare no conflict of interest.

Acknowledgements

This research was supported by the Wellcome Trust (4-year studentship WT096586 to CK, fellowships WT090051MA and 202831/Z/16/Z to MEW), the Royal Society (ET) and the BBSRC (project grant: BB/M024148/1). We thank Greg Daubney for performing histology; Richard Keithley for advice and assistance with chemometric analysis; and Katie Jennings, Paul Harrison, and Jaime McCutcheon for useful discussions about the data.

Abstract

Mechanisms for regulation of dopamine transmission are critical to its effects on behavior and vary by region. Recycling via the dopamine transporter (DAT) predominates in striatum, while degradation by catechol-O-methyltransferase (COMT) predominates in cortex. However, questions remain about whether and how each mechanism affects fast fluctuations in dopamine transmission in these regions and influences behavior. To address this issue, we used pharmacological blockade of each clearance mechanism to assess their roles in reward-guided decision making and in regulating sub-second dopamine transmission in striatum and cortex. We found that DAT and COMT selectively influence reward value updating in opposite directions, with DAT blockade impairing and COMT inhibition improving learning in a multi-step decision making task that requires mice to monitor changes in both optimal response strategy and reward probabilities. By contrast, neither drug influenced the speed of reversals following a change in action-state transition probabilities. In addition, DAT but not COMT influenced task engagement and motivation to work for reward in both the decision making task and a progressive ratio paradigm. Fast scan cyclic voltammetry recordings of evoked dopamine release in anesthetized mice revealed that DAT but not COMT blockade enhanced dopamine transients in nucleus accumbens. Unexpectedly, neither manipulation had an effect on evoked release in medial frontal cortex. Together, these data refine our understanding of how dopamine clearance mechanisms operate in different regions and at distinct timescales to shape aspects of reward-guided decision making.

Significance Statement

Dopamine transmission is tightly regulated by clearance mechanisms and these clearance mechanisms exhibit regional specificity. However, while we know a lot about how the activity of dopamine neurons relates to reward-guided behavior, the precise role of different clearance mechanisms in shaping these processes is much less understood. This is important, as dysfunctional clearance mechanisms have been implicated in many neuropsychiatric disorders. Here, we show specific and distinct roles for two clearance mechanisms – the dopamine transporter and catechol-O-methyltransferase – in reward value, but not action-state probability, updating during multi-step decision making, and in regulating striatal and cortical dopamine transients. Our findings demonstrate how regulation of dopamine transmission, over distinct timescales and in different brain regions, can influence multiple aspects of reward-guided behavior.

Introduction

Many studies have examined the relationship between the activity of dopamine neurons and behavior. However, clearance mechanisms, which are critical for the temporal and spatial regulation of dopamine's actions, have been much less well studied. Clearance mechanisms exhibit regional specificity within the mesocorticolimbic system. Recycling of dopamine via the dopamine transporter (DAT) predominates in nucleus accumbens (NAc) and other striatal regions (Sulzer et al., 2016). In medial frontal cortex (MFC), where DAT is sparse and located relatively distantly from dopamine terminals (Ciliax et al., 1999; Sesack et al., 1998), enzymatic degradation, particularly by catechol-*O*-methyltransferase (COMT), is a more prominent method of clearance (Karoum et al., 1994; Tunbridge et al., 2006).

Because of this regional specificity, DAT and COMT are often thought to contribute to different dopamine-mediated behaviors. DAT, through its regulation of striatal dopamine signaling (Budygin et al., 1999; Giros et al., 1996; Raevskii et al., 2002), influences reward-guided behavior, particularly its motivational components (Cagniard et al., 2006a, 2006b; Peciña et al., 2003; Zhuang et al., 2001). COMT, by contrast, contributes to regulation of dopamine levels in cortex (Käenmäki et al., 2010; Lapish et al., 2009; Slifstein et al., 2008; Tunbridge et al., 2004) and thus has been most studied in relation to cognitive functions influenced by cortical dopamine, including working memory, attention, and cognitive flexibility (Apud et al., 2006; Barkus et al., 2016; Colzato et al., 2010; Farrell et al., 2012; Tunbridge et al., 2004).

However, the function and effects of DAT and COMT are likely not so neatly separable. For instance, reward anticipation and receipt drive increases in dopamine levels in both striatum and frontal cortex (Bassareo et al., 2002, 2007; Ellwood et al., 2017), and COMT influences the

magnitude of this effect in frontal cortex (Lapish et al., 2009). In more complex settings, dopamine levels and COMT mediate the balance between model-free and model-based reinforcement learning systems (Doll et al., 2016; Sharp et al., 2016; Wunderlich et al., 2012). The circuitry involved in these behavioral effects is unclear because, although DAT predominates in striatum and COMT in cortex, both molecules are expressed to some degree in both regions (Laatikainen et al., 2013; Matsumoto et al., 2003; Sesack et al., 1998). Moreover, reciprocal interactions between cortical and striatal dopamine (Clarke et al., 2014; Kellendonk et al., 2006; Pycock et al., 1980) raise the possibility of indirect effects.

Thus, questions remain about whether the actions of DAT and COMT are truly separable. In the behavioral realm, it is unclear to what extent DAT affects learning about rewards as well as the motivation to pursue them, particularly during more complex decision making tasks that recruit both striatal and cortical regions. While some studies have suggested that individual variation in DAT levels is related to development of reward-based response biases (Kaiser et al., 2018), direct pharmacological or genetic manipulations of DAT do not necessarily cause evident changes in reward learning (Cagniard et al., 2006a, 2006b; Costa et al., 2014). The effects of COMT manipulations on reward-guided behavior, meanwhile, remain largely unexplored (Tunbridge et al., 2012). At the level of transmission, while fast and transient fluctuations in dopamine transmission are known to produce prediction error-like signals in both striatum and cortex (Ellwood et al., 2017; Hart et al., 2014), the extent to which DAT and COMT regulate this fast transmission – as opposed to slower minute-by-minute changes in dopamine levels – requires further study, particularly in cortex.

Here, we used behavioral testing, *in vivo* electrochemistry, and pharmacology to investigate the influence of DAT and COMT on reward-guided behavior and dopamine transmission in mice. Systemic administration of a DAT blocker (GBR-12909) or a COMT inhibitor (tolcapone) allowed

us to directly contrast the role of each clearance mechanism. We investigated how each agent influenced both motivation to work for reward and flexible multi-step decision making that required mice to adapt to changes in optimal response strategy and reward probabilities. Finally, to understand how DAT blockade and COMT inhibition affect the dynamics of fast fluctuations in dopamine, we recorded evoked dopamine transients in the NAc and MFC in anesthetized animals using fast scan cyclic voltammetry.

Materials & Methods

Animals

Male C57BL/6 mice were obtained from Envigo (formerly Harlan). Male mice were used in this study because COMT has previously been shown to exhibit sexually dimorphic effects (Harrison and Tunbridge, 2008). Mice were aged 9-26 weeks for behavioral experiments and 10-16 weeks for voltammetry experiments. Animals were housed on a 12/12-hr light/dark cycle; all behavioral tests were conducted during the light phase. Mice were food deprived to 85-90% of free feeding weight for the locomotor activity, progressive ratio, and multi-step decision making tasks, and were water deprived for 3hrs prior to test sessions for the sucrose preference test. Food and water were provided *ad libitum* in all other cases. All mice were habituated to handling – including the restraint position used during injections – before experiments began. Care and testing of all animals was conducted under the auspices of the UK Home Office laws and guidelines for the treatment of animals under scientific procedures and of the local ethical review board at the University of Oxford.

Drugs

Mice were administered the selective, brain-penetrant COMT inhibitor tolcapone (30mg/kg; TRC Inc) (Barkus et al., 2016; Männistö and Kaakkola, 1999) and/or the selective DAT blocker GBR-12909 dihydrochloride (Tocris) (Izenwasser et al., 1990; Rothman et al., 1989). GBR-12909 dihydrochloride was administered at 6mg/kg, a dose that in pilot experiments produced increased locomotor activity but not stereotypy, as assessed using criteria adapted from Creese and Iversen (1974) for use in mice (data not shown). D-amphetamine (in sulfate formulation; Tocris) was used at 4mg/kg (Avelar et al., 2013; Daberkow et al., 2013), and the NET blocker atomoxetine (in hydrochloride formulation; Tocris) at 1mg/kg (Bymaster et al., 2002; Koda et al., 2010). All drugs were dissolved in 20% hydroxypropyl-beta-cyclodextrin (Acros Organics) in 0.9% saline (AquPharm), which served as a vehicle control in all experiments. All drugs were delivered by intraperitoneal injection, with an injection volume of either 5mL/kg or 10mL/kg (multi-step decision making task only). Drug administration timings were designed to account for differences in drug time courses of action.

Behavioral tasks

Sucrose preference test

Sucrose preference was assessed in open-top cages equipped with two water bottles. Mice (previously run on locomotor/stereotypy test for drug dosage assessment and counterbalanced for prior drug exposure) were tested for 5hrs each day for a total of 7 days (Days 1-3: water exposure only; Days 4-7: sucrose exposure). Bottles were weighed immediately before and after testing to determine consumption. On day 7, mice received two injections – the first (tolcapone or vehicle) 1hr before testing and the second (GBR-12909 or vehicle) immediately

before testing. Preference for sucrose solution (10% weight/volume; Sigma Aldrich) was assessed as a ratio of sucrose consumption to total consumption.

Progressive ratio task

A total of 24 mice (two cohorts: n = 12 previously used for locomotor activity and sucrose preference tests, counterbalanced for prior drug exposure; n = 12 test naïve, given a vehicle injection 3 days prior to testing) were tested on the progressive ratio task. Of these, 3 mice were excluded from the analysis: 2 due to failed injections and 1 that was unable to learn the complete task. The timing of experimental drug administration differed between the two cohorts. Cohort 1 received the first injection 120min and the second injection 60min before the start of the session, whereas cohort 2 received the first injection 105min and the second injection 15min before the session. No notable effects of cohort were observed (data not shown), so findings from the two cohorts are reported together.

The task was conducted as previously described (Sharma et al., 2012) in standard operant chambers (Med Associates Inc). Rewards consisted of 60μL of 10% sucrose solution. Animals were trained on increasing fixed ratio (FR) schedules until they were able to earn ≥ 15 rewards and achieved an active:inactive lever press ratio of ≥ 3:1 on an FR5 schedule over 2 consecutive days. During progressive ratio (PR) sessions, the number of active lever presses required to obtain each subsequent reward was increased according to the following equation: number of required lever presses = $5*e^{i*0.16} – 5$ ('i': trial number). Drug effects on behavior were assessed by giving mice two systemic injections prior to PR test sessions: tolcapone or vehicle, followed by GBR-12909 or vehicle. Each mouse received all possible drug combinations over four PR test sessions according to a counterbalanced within-subjects design. Drug testing days

were interleaved with two washout days during which no injections were given: one day of testing on an FR5 schedule and one day of testing on a PR schedule.

Multi-step decision making task

The task was adapted from the two-step task developed by Daw et al. (2011) for dissociating model-based and model-free reinforcement learning in humans, as reported in Akam et al. (2015, 2017). A total of 16 mice began training on the task; after 6 days the 8 animals that had performed the most trials during that time were selected for continued training and subsequent drug testing.

The task was run in 8 custom made 12x12cm operant boxes controlled using pyControl (https://pycontrol.readthedocs.io). The behavioral apparatus consisted of 4 nose poke ports; a 'high' and a 'low' poke in the center flanked by 'left' and 'right' pokes (Figure 2A). Each trial started with the high and low pokes lighting up. The subject chose high or low, causing either the left or right poke to light up. The subject then poked the illuminated side for a probabilistic reward (20% weight/volume sucrose solution; Sigma Aldrich). At any point in time, one reward port had a high probability of giving reward (0.8) and the other a low reward probability (0.2). Similarly, a particular first-step action (high or low) usually led to a particular second-step state (left or right port active) ("common" transitions, 80% of trials), though sometimes led to the opposite state ("rare" transitions, 20% of trials).

Unlike other recent rodent adaptations of multi-step decision tasks (Groman et al., 2018; Hasz and Redish, 2018; Miller et al., 2017), both the reward probabilities in the second-step states and the transition probabilities linking the first-step actions to the second-step states reversed in blocks. Block transitions were triggered based on the subject's behavior, occurring 20 trials

after an exponential moving average (tau = 8 trials) of choices crossed a 75% correct threshold. Reversals in reward probability occurred twice as often as reversals in transition probability.

Subjects encountered the full trial structure from the first day of training. The only task parameters that were changed over the course of training were the state and reward transition probabilities and the reward sizes; reward size was gradually reduced and the reward and transition probabilities gradually adjusted over 38 days of training as mice became progressively more engaged with the task and learned to perform it better (see Table 1 for details). All animals had at least 12 sessions with the final task parameters prior to drug administration. Animals were considered fully trained and ready for pharmacological testing when the group average met the following criteria: (1) a 3-day average of >400 trials per session, (2) a 3-day average of >4 reversals per session, and (3) a 3-day average combined reversal learning speed of <30 trials.

Pharmacological manipulations were performed every second day of testing using a within-subjects design. On intervening days mice were run on the task but received no injection. Tolcapone or its vehicle control were administered 90min prior to the start of the session; GBR-12909 or its vehicle control were administered 15min prior to the start of the session. Subjects received a total of 8 of each drug and 5 of each vehicle injection, with order counterbalanced across animals.

Fast-scan cyclic voltammetry

A total of 100 mice were used for FCV recordings (NAc n = 43; MFC n = 57). Data from 57 (NAc n = 23; MFC n = 34) of these animals were included in the final analysis; group sizes for each region and drug treatment of the between-subjects design are shown in Table 2. Animals

were excluded due to death prior to the completion of the experiment (NAc n = 3; MFC n = 1), failed drug injection (NAc n = 6; MFC n = 4), or failure to meet data quality control criteria (NAc n = 9; MFC n = 13). Quality control criteria were (1) a ratio of standard deviation of pre-stimulation noise to peak height ≤ 0.5, (2) a detectable peak consistently occurring ≤ 3sec after stimulation, and (3) peak timing varying by ≤ 2sec across 30min bins. Finally, 2 NAc and 5 MFC animals were excluded due to poor fit ($R^2 < 0.5$) of the exponential decay function in the kinetics analysis.

Electrode fabrication and implantation

Recording and reference electrodes were made in-house (Papageorgiou et al., 2016; Syed et al., 2016) and pre-calibrated in a flow cell (Sinkala et al., 2012) to allow conversion of recorded signals from current (nA) into concentration (nM). When calibration factors were unavailable (9 electrodes), a mean calibration factor was used. The stimulating electrodes were untwisted, bipolar, stainless steel electrodes measuring 0.15mm in diameter (PlasticsOne).

Mice were implanted with electrodes under anesthesia (isoflurane induction [Abbott Laboratories Ltd] and urethane maintenance [Sigma Aldrich]). Mice were administered glycopyrronium bromide (0.01-0.02mg/kg i.p. alongside urethane; MercuryPharm Ltd) and glucosesaline (0.5% in 0.9% saline; ~7mL every 3hrs; Aquapharm), and temperature was maintained at 35.0-37.0°C with a rectal probe and heating blanket. Recording electrodes were placed either in the NAc core (+1.40 anterial-posterior [AP] and -0.75 medial-lateral [ML] from bregma; -3.50 to -4.75 dorsal-ventral [DV] from the skull) or prelimbic medial frontal cortex (+2.50 AP and -0.50 ML from bregma, -1.30 to -2.30 DV from skull). Stimulating electrodes targeted the posterior-medial portion of the VTA, where the cell bodies of dopamine neurons projecting to both NAc and MFC are located (-3.50 AP and -0.35 ML from bregma; -4.00 to -

4.80 DV from brain surface) (Lammel et al., 2008, 2011, 2014) . A reference electrode was placed in the contralateral hemisphere (+4.80 AP and +1.00 ML from bregma).

FCV recordings

Voltammetric recordings were made as previously described (Syed et al., 2016). Dopamine release was induced by passing current through the stimulating electrode via a DS3 Stimulator (Digitimer) (stimulation parameters: pulse number = 60 pulses, frequency = 50Hz, amplitude = 300µA, pulse width = 2ms, pulse phase = biphasic), based on previous literature (Yavich et al., 2007; Yorgason et al., 2011) and on pilot experiments that established the parameters required to reliably evoke detectable dopamine release in both the NAc and MFC. Note that as the cortex is innervated by significant noradrenergic as well as dopaminergic fibers (Lindvall et al., 1978; Slopsema et al., 1982), and because dopamine and noradrenaline have very similar cyclic voltammograms (Adams, 1976; Michael and Wightman, 1999), signals recorded in cortex using FCV can only be identified as catecholaminergic, not as definitively dopaminergic. However, given that previous studies have shown that electrical stimulation of the VTA predominantly evokes dopamine rather than noradrenaline release in MFC (Shnitko and Robinson, 2014), we will refer to such signals as 'dopamine' even though a contribution of noradrenaline cannot be entirely ruled out. Stimuli were generated and recordings collected using Tarheel CV (National Instruments).

Evoked signals decayed over time and so were allowed to stabilize (stimulating every 5-10mins for ~2.5hrs during NAc, and ~1hr during MFC, recordings). After the stabilization period stimulations were made every 5min. Following a 30min pre-drug baseline period, tolcapone or vehicle was administered. GBR-12909 or vehicle was then given after a further 90min of recording, and recordings continued for several hours. Amphetamine was tested in two groups

of animals: those that had received tolcapone or vehicle and naïve animals. There were no differences in signal decay between these groups (data not shown), so they were combined. Atomoxetine was only administered to drug naïve animals. Data were averaged into 15min bins for analysis: the effects of tolcapone were assessed at 85mins, of GBR-12909 and amphetamine at 30mins, and of atomoxetine at 60min post-administration. Once recording was complete, electrode placement was ascertained as previously described (Syed et al., 2016).

Experimental design and statistical analyses

Statistical analyses were conducted in SPSS versions 20 and 24 (IBM Computing), with the exception of the multi-step decision-making task (described further below), with significance set at $\alpha = 0.05$. With the exceptions noted below, data were analyzed using analysis of variance (ANOVA), with drug group(s) (and time, where relevant) as factors. For repeated-measures ANOVAs, Greenhouse-Geisser corrections were applied where data failed Mauchley's test of sphericity. Simple main effects analyses were conducted as necessary when significant interactions were found. Least square difference pairwise comparison tests were used to assess which groups were driving any significant main or interactive effects.

Progressive ratio task analysis

The main outcome measure was cumulative active lever presses over the session. We also examined cumulative inactive lever presses as a measure of general activity levels, the average reward collection latency, and the average re-engagement latency (the interval between the animal exiting the magazine after reward delivery and its next lever press). The significance of drug effects on these measures was assessed using repeated-measures ANOVAs, with drug 1 (tolcapone or vehicle) and drug 2 (GBR-12909 or vehicle) as within-subjects factors.

Multi-step decision making task analysis

For the multi-step decision making task, analysis of pharmacological manipulations was restricted to the first 90 minutes of each session. Except where stated otherwise drug effects were evaluated using repeated-measures ANOVAs. As the two different drugs each had a corresponding vehicle condition (see above), within subject factors were vehicle/drug – differentiating both vehicle from both drug conditions – and GBR-12909/tolcapone – differentiating GBR-12909 and its respective vehicle from tolcapone and its respective vehicle. Drug effects on trial rate (Figures 2D, 3A) were quantified as the mean number of trials each subject performed over the first 90 minutes. Drug effects on second-step reaction times (Figures 2E, 3B) were quantified using the median reaction time for each subject for each condition, with common/rare transition as an additional within-subject factor.

Drug effects on subjects' ability to track which option had highest reward probability were quantified by looking at the fraction of correct choices at the end of blocks and the speed of adapting to reversals (Figures 2F, 3C). The fraction of correct choices in the last 15 trials of each block was evaluated for each subject in drug and vehicle conditions and compared using a repeated-measures ANOVA. Two different approaches were used to quantify the speed of adapting to reversals. The fraction of correct choices in the first 15 trials of each block was calculated for each subject separately following reversals in the reward and transition probabilities and compared using a repeated-measures ANOVA with reversal type (reward or transition) and drug condition as within subject factors. To get a more fine-grained picture of how adaptation to reversals was affected by the drugs, the choice probability trajectory following reversals was fit by a sum of two exponential decays defined by the equation:

$$P_n = P_c - (P_c - P_0)(w_f\, e^{-\frac{n}{\tau_f}} + (1 - w_f)e^{-\frac{n}{\tau_s}})$$

Where P_n is the probability on trial n of choosing the option that was correct following the reversal, P_c is the asymptotic probability of choosing the correct option, defined as the cross-subject mean fraction of correct choices over the last 15 trials of all blocks, P_0 is the initial probability of choosing the correct option (calculated as $1 - P_c^i$, where P_c^i is the fraction of correct choices at the end of blocks preceding reversals of type being analyzed), τ_f is the time constant of the fast exponential decay, τ_s is the time constant of the slow exponential decay, and w_f is the weighting applied to the fast component relative to the slow component.

We used permutation testing to evaluate whether differences between drug and corresponding vehicle condition were significant, with the analysis performed independently for GBR-12909 and tolcapone. The curve was fit using a squared error cost function to the cross-subject mean choice probability trajectory for drug and vehicle conditions, and the difference $\Delta \mathrm{x}_{true}$ between drug and vehicle conditions was evaluated for parameters τ_f, τ_s, w_f. We then constructed an ensemble of 5000 permuted datasets in which the assignments of sessions to the drug and vehicle conditions were randomized. Randomization was performed within subjects, such that the number of sessions from each subject in each condition was preserved. For each permuted dataset we re-ran the analysis and evaluated the difference in each parameter between the two conditions, to give a distribution of $\Delta \mathrm{x}_{perm}$, which in the limit of many permutations is the distribution of $\Delta \mathrm{x}$ under the null hypothesis that there is no difference between the conditions. The two tailed p value for the observed difference is given by:

$$P = 2 \min\left(\frac{\mathrm{M}}{N}, 1 - \frac{\mathrm{M}}{N}\right)$$

Where N is the number of permutations and M is the number of permutations for which $\Delta \mathrm{x}_{perm} > \Delta \mathrm{x}_{true}$.

The statistical significance of trial-to-trial learning, and its modulation by drug treatment, was assessed using a logistic regression model. The model predicted repeating the previous choice as a function of trial outcome (rewarded or not), transition (common or rare), and their interaction. We additionally included two predictors capturing choice biases: one for a bias towards the high/low poke, and one for rotational bias – i.e. a tendency to choose high / low following trials that ended in the left / right second-step, which is observed in some animals on this task (Akam et al., 2017). We further included a predictor that promotes repeating correct choices. This prevents correlation between action values at the start of the trial and subsequent trial events from biasing on the transition-outcome interaction predictor loading (Akam et al., 2015).

The regression model was fit to a dataset comprising tolcapone, GBR-12909, and their corresponding vehicle sessions. Drug effects were modeled by interacting each predictor with *vehicle/drug* – differentiating both vehicle from both drug conditions, and *GBR-12909/tolcapone* – differentiating GBR-12909 and its respective vehicle from tolcapone. All coefficients were treated as independent random effects across subjects and the resulting hierarchical regression was fit using the lme4 mixed effects package (Bates et al., 2007) in the R statistical language (R Development Core Team, 2010). Random effects whose variance fit to zero were removed from the model to enable the fit to converge. This did not remove random effects for any predictors with significant fixed effects. P values were calculated using the LmerTest package (Kuznetsova et al., 2017) using Satterthwaite's method for approximating degrees of freedom. The regression model fit to the complete dataset indicated significant three way interactions between model predictors, vehicle/drug and GBR-12909/tolcapone. To unpack what was driving this interaction, we subsequently performed separate model fits for each drug with its corresponding vehicle, interacting the base predictors with vehicle/drug condition.

We also explored whether fitting reinforcement learning (RL) models to the multi-step decision making task data could provide insight into the behavioral strategies used by the mice and how the drugs affected these. We first modeled behavioral data from baseline sessions (when no injections were given). To do this, we compared a set of models generated by adding or removing single features from the model found to best describe behavior on this task in Akam et al., 2017. These features included: forgetting about the values and state transitions for not-chosen actions, action perseveration effects spanning multiple trials, and representation of actions both at the level of the choice they represent (e.g. high poke) and the motor action they require (e.g. left → high movement) (for full details see Akam et al., 2017). Models were compared using the integrated Bayes Information Criterion (BIC) score. In addition to modeling baseline session data, we compared the signed difference between maximum likelihood parameter estimates after administration of GBR-12909 or tolcapone with their respective vehicles. However, although some changes following drug administration were found in the reinforcement learning model analysis, the complexity of the model meant that no effects survived multiple comparison correction for the number of model parameters. Therefore, the results of the RL modeling analysis on drug session data are not presented.

FCV recording analysis

FCV recordings were processed using software written in LabVIEW and custom Matlab scripts. Dopamine levels were extracted using a chemometric approach based on training sets from individual animals (Heien et al., 2005; Keithley et al., 2009, 2010). Cyclic voltammograms were low-pass filtered at 2kHz and background subtracted using the 5 scans prior to stimulation.

Drug effects on evoked dopamine were assessed by quantifying several features of the evoked transients, including: the peak height; the latency from the start of stimulation to the peak; and the rate of decay of the signal from the peak to T50 (the time when the signal had decayed to half the peak height) (Figure 4B). (In cases where the signal did not fall to half its peak height, the decay over the 3sec following the peak was used.) Prior to statistical analysis, parameters were normalized to the pre-drug baseline signal and binned across three individual recordings at the time of interest. The significance of drug effects on each parameter was determined using repeated-measures ANOVAs, with time as the within-subjects factor and drug 1 (tolcapone or vehicle) and drug 2 (GBR-12909 or vehicle) as between-subjects factors. As we found no interactive effects of the two experimental drugs, the effects of COMT inhibition and DAT blockade are presented separately.

Results

DAT blockade, but not COMT inhibition, increases motivation to work for reward

Before beginning behavioral testing, we assessed the effects of DAT blockade and COMT inhibition on basic reward processing using a sucrose preference test. We found no influence of either tolcapone or GBR-12909 on either sucrose preference or absolute sucrose or water consumption (all $F < 1.7$, $p > 0.21$, univariate ANOVAs). These results indicate that drug effects on the hedonic properties of sucrose rewards did not influence subsequent behavioral experiments, although it is possible that the sucrose preference test was not sensitive enough to rule out such confounding effects.

We next investigated the influence of DAT blockade and COMT inhibition on reward guided behavior using a progressive ratio (PR) task. DAT blockade selectively increased active lever presses (main effect of drug 2: $F(1,19) = 26.4$, $p = 6x10^{-5}$, $\eta_p^2 = 0.582$, repeated-measures ANOVA) without affecting inactive lever presses (main effect of drug 2: $F(1,19) = 2.7$, $p = 0.118$, repeated-measures ANOVA) (Figure 1C,D). DAT blockade also produced a speeding of re-engagement latencies (main effect of drug 2: $F(1,19) = 15.4$, $p = 0.001$, $\eta_p^2 = 0.447$, repeated-measures ANOVA) and a trend-level effect on reward collection latency (main effect of drug 2: $F(1,19) = 4.1$, $p = 0.056$, $\eta_p^2 = 0.179$, repeated-measures ANOVA) (Figure 1E,F). In contrast, there were no reliable effects of COMT inhibition, either in isolation or in interaction with DAT blockade, on any measures in the PR task (all $F < 3.1$, $p > 0.09$, repeated-measures ANOVAs) (Figure 1C-F).

Both DAT blockade and COMT inhibition modulate value updating during multi-step decision making

To assess how regulation of dopamine transmission affects flexible reward-guided decision making, we used a multi-step decision task in which mice had to learn which of two options to select (high / low nose pokes) to gain access to a high probability reward port (left / right ports) (Figure 2A). Maximizing the reward rate on the task requires choosing the action at the first step that commonly leads to the second step state with high reward probability, and tracking this correct action across reversals in the reward and transition probabilities.

As can be seen in Figure 2c, mice learned to do this proficiently. By the end of training, animals were performing 442.2 ± 28.8 trials and 5.7 ± 0.6 blocks per session, completing 70.5 ± 5.2 trials per block, and obtaining reward on 51.6 ± 0.5 percent of trials (mean ± SEM across animals over the 3 days before the first injection). We fit reinforcement learning models using

data from baseline (no drug or vehicle injections) sessions. The best fitting model employed a mixture of model-based and model-free control, along with several additional factors that captured features of the animals' behavior such as choice and motor perseveration, motor biases, forgetting of non-chosen values, and state-action transitions (Figure 2-1), in line with previous reports (Akam et al., 2017).

We first assessed motivational effects of DAT blockade and COMT inhibition by evaluating how they affected the speed with which subjects performed the task. DAT blockade but not COMT inhibition increased the rate at which subjects performed trials (main effect of veh/drug: $F(1,7) = 9.79$, $p = 0.016$, veh/drug x DAT/COMT interaction: $F(1,7) = 6.52$, $p = 0.038$, post-hoc paired t-tests, DAT: $t(7) = 3.34$, $p = 0.012$, COMT: $t(7) = 0.023$, $p = 0.98$) (Figure 2D, 3A). There was also a selective speeding of second-step reaction times following DAT blockade (i.e. the latency from choosing high or low to entering the active side poke) (main effect of veh/drug: $F(1,7) = 8.20$, $p = 0.024$, veh/drug x DAT/COMT interaction: $F(1,7) = 7.76$, $p = 0.027$) (Figure 2E, 3B). This decrease in reaction time was larger following rare than common transitions after drug administration (veh/drug x common/rare interaction: $F(1,7) = 9.26$, $p = 0.019$), though the 3-way interaction did not quite reach significance (veh/drug x common/rare x DAT/COMT interaction: $F(1,7) = 4.35$, $p = 0.076$).

We next assessed how DAT blockade and COMT inhibition affected subjects' ability to track the correct option across reversals in the second-step reward probabilities and action-state transition probabilities. Neither drug affected the probability of choosing the correct (i.e. reward maximizing) option at the end of blocks (repeated-measures ANOVA: all $F(1,7) < 1.59$, $p > 0.24$) (Figure 2F, 3C). However, DAT blockade and COMT inhibition differentially affected how rapidly subjects adapted to reversals, as assessed by the fraction of correct choices in the first 15 trials following reversals in the reward or transition probabilities (repeated-measures ANOVA:

veh/drug x DAT/COMT interaction $F(1,7) = 5.66$, $p = 0.049$, veh/drug x DAT/COMT x reward/transition-reversal interaction $F(1,7) = 6.88$, $p = 0.034$). To better understand the nature of each drug's effect on reversals of each type, we therefore separately analyzed the effects of each drug on performance following block switches. As can be observed in Figure 2F, DAT blockade selectively impaired performance following reversals in the reward probabilities but had no influence after reversals in transition probabilities (main effect of DAT: $F(1,7) = 17.4$, $p = 0.0042$, DAT x reward/transition interaction: $F(1,7) = 5.81$, $p = 0.046$). COMT inhibition did not reliably change reversal performance (repeated-measures ANOVA: main effect of COMT: $F(1,7) = 0.57$, $p = 0.47$, DAT x reward/transition interaction $F(1,7) = 1.65$, $p = 0.24$).

To obtain a more fine-grained picture of drug effects on reversal behavior we fit a double exponential curve to the choice probability trajectories following reversals in reward and transition probabilities. We then used a permutation test for each drug and reversal type independently to assess whether the fits to each reversal type were changed by the drug manipulations. DAT blockade again significantly increased the time constant of adaptation to reversals in the reward probabilities ($p = 0.0012$, permutation test, Table 3) but did not affect how quickly subjects adapted to reversals in the transition probabilities (permutation test $p > 0.23$ for all fit parameters, Table 3) (Figure 2F). Therefore, while normal DAT function is important for updating of action values, it has limited influence on updating of action-state transition probabilities. In addition, this curve fitting analysis indicated that COMT inhibition reduced the time constant for adaptation to reversals in the second-step state reward probabilities ($p = 0.049$, permutation test, Table 4) (Figure 3C). This effect was again specific to reward probability reversals, as COMT inhibition had no effect on how fast subjects adapted to reversals in the transition probabilities (permutation test $p > 0.3$ for all fit parameters, Table 4).

Analysis of stay probabilities indicated that, consistent with our previous work with this task (Akam et al., 2017), both rewarded outcomes and common transitions promoted repeating choices (Figure 2G, 3D, $P < 0.001$, mixed effects logistic regression), but the transition-outcome interaction did not ($P = 0.78$). In online learning of action-state transition probabilities, a model-based reinforcement learning strategy tends to generate a main effect of transition rather than a transition-outcome interaction (Akam et al., 2015, 2017). Therefore, the regression analysis supports the RL modeling described earlier in demonstrating that the behavior of the mice involved a model-based component and was not simply driven by model-free reinforcement learning.

However, when we assessed drug effects, we found that neither DAT nor COMT manipulations modified the influence of previous trials on the subsequent choice ($P > 0.086$ for all interactions of outcome and transition with drug condition). Instead, the drug manipulations affected the regression model's 'correct' predictor, which promotes repeating choices to the high reward probability option (correct x veh/drug x DAT/COMT interaction $P = 0.016$). To understand what was driving this three-way interaction, we separately analyzed stay probabilities for each drug with its corresponding vehicle. This indicated a trend towards a reduced tendency to choose the correct option only under DAT blockade ($p = 0.056$, mixed effects logistic regression) (COMT inhibition: $p = 0.28$). Therefore, although the immediate effect of trial events on the subsequent choice was intact under DAT blockade, cumulative learning across multiple trials, as indicated by the analysis of reversal performance, was impaired.

<u>DAT blockade, but not COMT inhibition, selectively influences evoked dopamine transients in NAc but not MFC</u>

To better understand the dynamics of dopamine transmission underlying the behavioral effects that we observed in the progressive ratio and multi-step decision making tasks, we studied the impact of DAT blockade and COMT inhibition on VTA-stimulation-evoked dopamine release in both striatum and frontal cortex in anesthetized animals. We were particularly interested in the influence of DAT and COMT on fast fluctuations in transmission. The timing and size of such fast transients are critical for precise reinforcement learning, but most previous studies have assessed the roles of DAT and COMT only at slower timescales (Carboni et al., 2006; Huotari et al., 1999; Lapish et al., 2009; Raevskii et al., 2002; Tammimaki et al., 2016).

DAT blockade enhanced the size of evoked dopamine release in the NAc core (Figure 4F,H). There was an interactive effect of the DAT blocker and time on signal peak height ($F(1,19) = 18.1$, $p = 0.0004$, $\eta_p^2 = 0.488$, repeated-measures ANOVA) as well as main effects of both the DAT blocker ($F(1,19) = 7.8$, $p = 0.011$, $\eta_p^2 = 0.292$) and time ($F(1,19) = 11.5$, $p = 0.003$, $\eta_p^2 = 0.376$) on this measure. Post-hoc tests confirmed an increase in evoked release selectively following administration of the DAT blocker but not of vehicle ($p = 0.002$). DAT blockade also affected the kinetics of evoked NAc transients: blockade slowed both the latency to peak and the decay from peak (Figure 4J). There was an interactive effect of the DAT blocker and time on the latency to peak ($F(1,19) = 12.6$, $p = 0.002$, $\eta_p^2 = 0.399$, repeated-measures ANOVA), along with main effects of both the DAT blocker ($F(1,19) = 7.8$, $p = 0.011$, $\eta_p^2 = 0.293$) and time ($F(1,19) = 6.3$, $p = 0.021$, $\eta_p^2 = 0.249$) on this measure. Post-hoc tests again confirmed the selective effect of the DAT blocker on the latency to peak ($p = 0.002$). Furthermore, there was an interactive effect of the DAT blocker and time on the rate of decay from the peak ($F(1,19) = 6.5$, $p = 0.020$, $\eta_p^2 = 0.254$), although neither the main effect of DAT blockade nor that of time was significant (all $F < 2.9$, $p > 0.10$). Post-hoc tests showed a trend level effect of DAT blockade on post-peak kinetics compared to vehicle ($p = 0.078$) and a difference in the rate of

decay in the DAT blockade group before and after receiving the drug ($p = 0.007$) that was not seen in the vehicle group ($p = 0.565$).

In contrast, COMT inhibition had no effect on any index of evoked dopamine transmission in NAc core (Figure 4E,G,I). There were no main or interactive effects involving the COMT inhibitor on peak height, latency to peak, or decay from peak (all $F < 2.4$, $p > 0.14$).

Given the prominent role of COMT degradation in regulating cortical dopamine transmission, and the potential involvement of cortical as well as striatal regions in behavioral tests such as the multi-step decision making task, we followed up our recordings of VTA-stimulation evoked dopamine release in the NAc with similar recordings in the prelimbic MFC (Figure 5A). Evoked release in the MFC was, in general, considerably smaller than in the NAc (range of signal sizes at the start of the pre-drug baseline recording period, mean ± standard deviation: NAc = 11.48 ± 12.36 nA, MFC = 2.26 ± 0.96 nA). Nevertheless, we were able to record clear signals in our cortical experiments (Figure 5B).

Unexpectedly, we found no influence of either DAT blockade (Figure 5D,F,H) or COMT inhibition (Figure 5C,E,G) on evoked dopamine in prelimbic MFC. There were no main or interactive effects on any measure of evoked dopamine in MFC (all $F < 3.7$, $p > 0.07$) other than an interaction between the second drug injection and time on post-peak signal decay ($F(1,21) = 6.0$, $p = 0.023$, $\eta_p^2 = 0.222$, repeated-measures ANOVA on decay constant). This resulted from a difference between the vehicle and DAT blocker groups at the first time point (i.e. prior to the GBR-12909 injection) ($p = 0.018$) and is thus likely due to variability in the MFC decay constant data arising from the difficulty of performing an exponential fit on small cortical transients.

Given the lack of effect of either DAT blockade or COMT inhibition on evoked dopamine in the MFC, we tested the effect of amphetamine to ensure that we could detect drug-induced changes (Figure 5I). Amphetamine both increased the peak height ($t(15) = -4.6$, $p = 0.0003$, Hedges' $g = 2.279$, independent samples t-test) and decreased the decay constant value ($t(15) = 2.4$, $p = 0.030$, Hedges' $g = 1.180$, independent samples t-test) for evoked transients compared to vehicle. Similarly, blockade of the noradrenaline transporter (NET) using atomoxetine slowed clearance of cortical transients compared to vehicle controls ($t(13) = 2.6$, $p = 0.023$, Hedges' $g = 1.470$, independent samples t-test), although in this case without affecting transient peak height ($t(13) = 0.1$, $p = 0.915$, Hedges' $g = 0.060$, independent samples t-test) (Figure 5-2). Our data thus indicate that neither DAT nor COMT participates in the clearance of fast fluctuations in catecholamines in frontal cortex.

Discussion

Here we demonstrate distinct roles for DAT and COMT in reward-guided behavior. DAT blockade affected multiple facets of reward-guided behavior: it increased motivational drive in the PR task, stimulated task engagement during the multi-step decision making task, and additionally selectively impaired updating of reward values during the latter task. In contrast, COMT inhibition did not alter PR behavior but did improve reward value updating in the decision making task. FCV recordings confirmed a role for DAT recycling, but not COMT degradation, in regulating fast fluctuations in dopamine transmission in NAc. Unexpectedly, neither DAT blockade nor COMT inhibition affected evoked dopamine transients in prelimbic MFC, indicating that clearance mechanisms other than DAT and COMT contribute to regulation of cortical dopamine at sub-second timescales.

The significance of DAT for reward-guided learning and motivation

Our findings demonstrate that DAT influences several distinct aspects of reward-guided behavior. In the PR task, DAT blockade increased active lever presses and speeded task reengagement latency, consistent with evidence demonstrating its importance for motivational drive and the exertion of effort to obtain reward (Cagniard et al., 2006a, 2006b; Young and Geyer, 2010; Zhuang et al., 2001). However, the nature of the PR task makes it difficult to determine whether differences in learning might also contribute to these behavioral effects.

We therefore also employed a more complex multi-step sequential decision making task as a means of disentangling this confound. Unlike other similar multi-step paradigms in rodents (Groman et al., 2018; Hasz and Redish, 2018; Miller et al., 2017), the version we used included reversals in both reward and action-state transition probabilities. This not only reduced the chance for the animals to depend on a sophisticated habit-like strategy (Akam et al., 2015) but also allowed us to examine the influence of DAT and COMT on different aspects of behavioral flexibility. Consistent with previous work using the same task (Akam et al., 2017), the mice were sensitive to the transition structure of the task and exhibited behavior consistent with them using a mixture of model-based and model-free reinforcement learning.

Our findings suggest that DAT is important for multiple features of reward-guided behavior. In addition to increased trial rate and speeded responding following DAT blockade, we found a selective effect on reward updating: specifically, DAT blockade decreased a subjects' ability to adapt following reversals in second-step reward probabilities. This was not caused by differences in performance on and off the drug prior to reversals. Moreover, as the task included reversals in transitions as well as reward probabilities, we were able to demonstrate

that, strikingly, the effect of DAT blockade on learning was not observed following reversals in the transition probabilities linking first-step actions to second-step states, even though these also provided an opportunity for animals to adapt their behavior (in this case, at the first-step choice) in order to maximize rewards obtained. Therefore, the deficit cannot be attributed to a general behavioral inflexibility. Instead, our data are consistent with DAT influencing rapid reward-driven alternations in behavioral strategies and motivational components of reward-guided behavior.

While numerous studies have implicated DAT in the regulation of motivation, there is limited evidence linking it to reward learning, with many studies finding no clear effects of genetic or pharmacological disruption of DAT function on acquisition of either instrumental or Pavlovian associations (Cagniard et al., 2006a, 2006b; Costa et al., 2014; Kaiser et al., 2018; Peciña et al., 2003; Yin et al., 2006). Fast fluctuations in striatal dopamine correlate with reward prediction error signals, which are strongly linked to animals' ability to form certain reward-related associations (Day et al., 2007; Flagel et al., 2011; Pessiglione et al., 2006; Saddoris et al., 2015). In agreement with previous studies (Budygin et al., 1999; Huotari et al., 1999, 2002; Nomikos et al., 1990; Raevskii et al., 2002), we found that DAT blockade both increased and extended evoked NAc dopamine transients. Theoretically, these extended striatal dopamine release events could promote reinforcement learning by boosting reward prediction error-like signals. Alternatively, an increase in the duration of such transients could corrupt the precision of this encoding, reducing reinforcement learning efficiency. It is therefore likely that the behavioral consequences of DAT blockade on reward learning will depend on the paradigm. For instance, impairments might be more likely in paradigms like our multi-step decision making task – where trials are frequent, associations are changeable, and rewards are uncertain – because performing adaptively on such tasks is facilitated if subjects use precise contingent

learning of choice-state-outcome associations rather than non-contingent approximations based on recent choice or reward histories (Walton et al., 2011).

Our voltammetry recordings came from anesthetized animals and used supraphysiological stimulation parameters to compare release in striatum with that in cortex (where the lower density of dopamine terminals reduces signal-to-noise compared to striatum). Nevertheless, the effects of DAT blockade that we observed are similar to those seen on spontaneous dopamine transients in freely-moving animals following administration of nomifensine, a catecholamine transporter blocker, or stimulant drugs such as amphetamine (at least at low to moderate doses) (Daberkow et al., 2013; Robinson and Wightman, 2004).

We found no effect of DAT blockade on evoked release in prelimbic MFC, consistent with the sparse DAT content in this region (Sesack et al., 1998). DAT might still regulate cortical dopamine levels over longer timescales: some previous studies of DAT blockade have observed an effect on cortical dopamine levels measured with microdialysis over minutes (Carboni et al., 2006; Cass and Gerhardt, 1995; Tanda et al., 1997; Valentini et al., 2004), although evidence is mixed (Mazei et al., 2002; Pozzi et al., 1994; Weikop et al., 2007). Nonetheless, our data indicate that DAT recycling only plays a role in regulating fast dopamine transmission in striatum.

The role of COMT degradation

Research on COMT's role in behavior has largely focused on cognitive functions, while its significance for reward-guided behavior remains relatively unstudied (Tunbridge et al., 2012). Here, we tested the effects of inhibiting COMT in two reward-guided tasks. COMT inhibition had no effect on motivational aspects of behavior, in line with its absence of effect on NAc dopamine transmission (Acquas et al., 1992; Budygin et al., 1999; Garris and Wightman, 1995).

Nonetheless, there was a selective effect of COMT inhibition on value updating in the multi-step sequential decision making task. In contrast to DAT blockade, COMT inhibition speeded reversals. Although acute pharmacological inhibition of COMT is different than the chronic changes in enzymatic activity arising from the human COMT Val/Met polymorphism, this finding is concordant with reports of faster reinforcement learning in Met allele carriers, who have lower COMT activity than Val allele homozygotes, statistically significant by meta-analysis (Corral-Frías et al., 2016). Moreover, our data align with the proposal by Frank and colleagues that the behavior of Met allele carriers is more sensitive to single instances of negative feedback (Frank et al., 2007).

Given that COMT inhibition had little to no effect on the size or kinetics of evoked dopamine transients in either NAc or prelimbic MFC, COMT appears not to shape sub-second reinforcement signals. This might at first appear surprising given COMT's well-established role in regulating cortical dopamine transmission (Gogos et al., 1998; Käenmäki et al., 2010; Slifstein et al., 2008; Tunbridge et al., 2004). Indeed, the only previous study that investigated the influence of COMT on fast catecholamine transmission reported an increase in dopamine overflow compared to wild-type controls (Yavich et al., 2007). However, there are a number of important differences between this previous study and ours, notably, the method of COMT manipulation (a constitutive knock-out versus an acute pharmacological challenge) and the analysis approach (amperometric currents combined with periodic cyclic voltammograms versus principal component regression). The latter may be particularly important given the difficulty of separating dopamine from other potential chemical contaminants in cortex.

In addition, COMT's regulation of cortical dopamine transmission is complex and context-dependent: effects of COMT on cortical dopamine are typically only observed under conditions of potentiated dopamine transmission (Lapish et al., 2009; Tammimaki et al., 2016; Tunbridge et

al., 2004). While the precise synaptic location of COMT is not fully determined, it is likely situated on postsynaptic membranes or even inside postsynaptic neurons, possibly extrasynaptically (Chen et al., 2011; Myöhänen et al., 2010). This would limit its ability to directly modulate fast fluctuations in dopamine. Furthermore, given the lack of autoreceptors in cortically-projecting dopamine neurons (Gainetdinov and Caron, 2003; Lammel et al., 2008), there may also be less scope in cortex for indirect effects of tonic dopamine levels on evoked transients, as occurs in striatum (Grace, 1991; Sulzer et al., 2016). Instead, our data suggest that clearance mechanisms other than DAT and COMT regulate the kinetics of shorter-lived dopamine transients in the cortex. While FCV currently lacks chemical selectivity to separate dopamine from noradrenaline, the key point is that neither DAT nor COMT appears to be a major regulator of cortical catecholamine levels at the type of fast timescales required for precise reinforcement learning.

In conclusion, we demonstrate that both DAT and COMT regulate specific and distinct aspects of reward-guided behavior, although they had little influence on the balance of reinforcement learning strategies. While DAT regulates fast fluctuations of dopamine in NAc, these fluctuations are unaffected by both DAT and COMT in MFC. Taken together, our findings demonstrate the complex role of dopamine in multiple aspects of reward-guided learning, which appears to operate over distinct timescales and in different brain regions to mediate its effects.

References

Acquas, E., Carboni, E., Ree, R.H.A., Prada, M., and Chiara, G. (1992). Extracellular Concentrations of Dopamine and Metabolites in the Rat Caudate After Oral Administration of a Novel Catechol-O-Methyltransferase Inhibitor Ro 40–7592. J. Neurochem. *59*, 326–330.

Adams, R.N. (1976). Probing brain chemistry with electroanalytical techniques. Anal. Chem. *48*, 1126A–1138A.

Akam, T., Costa, R., and Dayan, P. (2015). Simple Plans or Sophisticated Habits? State, Transition and Learning Interactions in the Two-Step Task. PLOS Comput. Biol. *11*, e1004648.

Akam, T., Rodrigues-Vaz, I., Zhang, X., Pereira, M., Oliveira, R., Dayan, P., and Costa, R.M. (2017). Single-Trial Inhibition of Anterior Cingulate Disrupts Model-based Reinforcement Learning in a Two-step Decision Task. bioRxiv. doi: https://doi.org/10.1101/126292

Apud, J.A., Mattay, V., Chen, J., Kolachana, B.S., Callicott, J.H., Rasetti, R., Alce, G., Iudicello, J.E., Akbar, N., Egan, M.F., et al. (2006). Tolcapone Improves Cognition and Cortical Information Processing in Normal Human Subjects. Neuropsychopharmacology *32*, 1011–1020.

Avelar, A.J., Juliano, S.A., and Garris, P.A. (2013). Amphetamine augments vesicular dopamine release in the dorsal and ventral striatum through different mechanisms. J. Neurochem. *125*, 373–385.

Barkus, C., Korn, C., Stumpenhorst, K., Laatikainen, L.M., Ballard, D., Lee, S., Sharp, T., Harrison, P.J., Bannerman, D.M., Weinberger, D.R., et al. (2016). Genotype-Dependent Effects of COMT Inhibition on Cognitive Function in a Highly Specific, Novel Mouse Model of Altered COMT Activity. Neuropsychopharmacology.

Bassareo, V., Luca, M.A.D., and Chiara, G.D. (2002). Differential Expression of Motivational Stimulus Properties by Dopamine in Nucleus Accumbens Shell versus Core and Prefrontal Cortex. J. Neurosci. *22*, 4709–4719.

Bassareo, V., De Luca, M.A., and Di Chiara, G. (2007). Differential impact of pavlovian drug conditioned stimuli on in vivo dopamine transmission in the rat accumbens shell and core and in the prefrontal cortex. Psychopharmacology (Berl.) *191*, 689–703.

Bates, D., Sarkar, D., Bates, M.D., and Matrix, L. (2007). The lme4 package. R Package Version *2*, 74.

Budygin, E.A., Gainetdinov, R.R., Kilpatrick, M.R., Rayevsky, K.S., Männistö, P.T., and Wightman, R.M. (1999). Effect of tolcapone, a catechol-O-methyltransferase inhibitor, on striatal dopaminergic transmission during blockade of dopamine uptake. Eur. J. Pharmacol. *370*, 125–131.

Bymaster, F.P., Katner, J.S., Nelson, D.L., Hemrick-Luecke, S.K., Threlkeld, P.G., Heiligenstein, J.H., Morin, S.M., Gehlert, D.R., and Perry, K.W. (2002). Atomoxetine increases extracellular levels of norepinephrine and dopamine in prefrontal cortex of rat: a potential mechanism for efficacy in attention deficit/hyperactivity disorder. Neuropsychopharmacology *27*, 699–711.

Cagniard, B., Beeler, J.A., Britt, J.P., McGehee, D.S., Marinelli, M., and Zhuang, X. (2006a). Dopamine Scales Performance in the Absence of New Learning. Neuron *51*, 541–547.

Cagniard, B., Balsam, P.D., Brunner, D., and Zhuang, X. (2006b). Mice with chronically elevated dopamine exhibit enhanced motivation, but not learning, for a food reward. Neuropsychopharmacology *31*, 1362–1370.

Carboni, E., Silvagni, A., Vacca, C., and Di Chiara, G. (2006). Cumulative effect of norepinephrine and dopamine carrier blockade on extracellular dopamine increase in the nucleus accumbens shell, bed nucleus of stria terminalis and prefrontal cortex. J. Neurochem. *96*, 473–481.

Cass, W.A., and Gerhardt, G.A. (1995). In Vivo Assessment of Dopamine Uptake in Rat Medial Prefrontal Cortex: Comparison with Dorsal Striatum and Nucleus Accumbens. J. Neurochem. *65*, 201–207.

Chen, J., Song, J., Yuan, P., Tian, Q., Ji, Y., Ren-Patterson, R., Liu, G., Sei, Y., and Weinberger, D.R. (2011). Orientation and Cellular Distribution of Membrane-bound Catechol-O-methyltransferase in Cortical Neurons IMPLICATIONS FOR DRUG DEVELOPMENT. J. Biol. Chem. *286*, 34752–34760.

Ciliax, B.J., Drash, G.W., Staley, J.K., Haber, S., Mobley, C.J., Miller, G.W., Mufson, E.J., Mash, D.C., and Levey, A.I. (1999). Immunocytochemical localization of the dopamine transporter in human brain. J. Comp. Neurol. *409*, 38–56.

Clarke, H.F., Cardinal, R.N., Rygula, R., Hong, Y.T., Fryer, T.D., Sawiak, S.J., Ferrari, V., Cockcroft, G., Aigbirhio, F.I., Robbins, T.W., et al. (2014). Orbitofrontal Dopamine Depletion Upregulates Caudate Dopamine and Alters Behavior via Changes in Reinforcement Sensitivity. J. Neurosci. *34*, 7663–7676.

Colzato, L.S., Waszak, F., Nieuwenhuis, S., Posthuma, D., and Hommel, B. (2010). The flexible mind is associated with the catechol-O-methyltransferase (COMT) Val158Met polymorphism: Evidence for a role of dopamine in the control of task-switching. Neuropsychologia *48*, 2764–2768.

Corral-Frías, N.S., Pizzagalli, D.A., Carré, J.M., Michalski, L.J., Nikolova, Y.S., Perlis, R.H., Fagerness, J., Lee, M.R., Conley, E.D., Lancaster, T.M., et al. (2016). COMT Val158Met genotype is associated with reward learning: a replication study and meta-analysis. Genes Brain Behav. *15*, 503–513.

Costa, V.D., Tran, V.L., Turchi, J., and Averbeck, B.B. (2014). Dopamine modulates novelty seeking behavior during decision making. Behav. Neurosci. *128*, 556–566.

Creese, I., and Iversen, S.D. (1974). The role of forebrain dopamine systems in amphetamine induced stereotyped behavior in the rat. Psychopharmacologia *39*, 345–357.

Daberkow, D.P., Brown, H.D., Bunner, K.D., Kraniotis, S.A., Doellman, M.A., Ragozzino, M.E., Garris, P.A., and Roitman, M.F. (2013). Amphetamine Paradoxically Augments Exocytotic Dopamine Release and Phasic Dopamine Signals. J. Neurosci. *33*, 452–463.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. Neuron *69*, 1204–1215.

Day, J.J., Roitman, M.F., Wightman, R.M., and Carelli, R.M. (2007). Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. Nat. Neurosci. *10*, 1020–1028.

Doll, B.B., Bath, K.G., Daw, N.D., and Frank, M.J. (2016). Variability in Dopamine Genes Dissociates Model-Based and Model-Free Reinforcement Learning. J. Neurosci. *36*, 1211–1222.

Ellwood, I.T., Patel, T., Wadia, V., Lee, A.T., Liptak, A.T., Bender, K.J., and Sohal, V.S. (2017). Tonic or Phasic Stimulation of Dopaminergic Projections to Prefrontal Cortex Causes Mice to Maintain or Deviate from Previously Learned Behavioral Strategies. J. Neurosci. *37*, 8315–8329.

Farrell, S.M., Tunbridge, E.M., Braeutigam, S., and Harrison, P.J. (2012). COMT Val158Met Genotype Determines the Direction of Cognitive Effects Produced by Catechol-O-Methyltransferase Inhibition. Biol. Psychiatry *71*, 538–544.

Flagel, S.B., Clark, J.J., Robinson, T.E., Mayo, L., Czuj, A., Willuhn, I., Akers, C.A., Clinton, S.M., Phillips, P.E.M., and Akil, H. (2011). A selective role for dopamine in stimulus–reward learning. Nature *469*, 53–57.

Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T., and Hutchison, K.E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc. Natl. Acad. Sci. *104*, 16311–16316.

Gainetdinov, R.R., and Caron, and M.G. (2003). Monoamine Transporters: From Genes to Behavior. Annu. Rev. Pharmacol. Toxicol. *43*, 261–284.

Garris, P.A., and Wightman, R.M. (1995). Distinct pharmacological regulation of evoked dopamine efflux in the amygdala and striatum of the rat in vivo. Synapse *20*, 269–279.

Giros, B., Jaber, M., Jones, S.R., Wightman, R.M., and Caron, M.G. (1996). Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. Nature *379*, 606–612.

Gogos, J.A., Morgan, M., Luine, V., Santha, M., Ogawa, S., Pfaff, D., and Karayiorgou, M. (1998). Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. Proc. Natl. Acad. Sci. *95*, 9991–9996.

Grace, A.A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: A hypothesis for the etiology of schizophrenia. Neuroscience *41*, 1–24.

Groman, S.M., Massi, B., Mathias, S.R., Curry, D.W., Lee, D., and Taylor, J.R. (2018). Neurochemical and behavioral dissections of decision-making in a rodent multi-stage task. J. Neurosci. 2219–18.

Harrison, P.J., and Tunbridge, E.M. (2008). Catechol-O-methyltransferase (COMT): a gene contributing to sex differences in brain function, and to sexual dimorphism in the predisposition to psychiatric disorders. Neuropsychopharmacology *33*, 3037–3045.

Hart, A.S., Rutledge, R.B., Glimcher, P.W., and Phillips, P.E.M. (2014). Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term. J. Neurosci. *34*, 698–704.

Hasz, B.M., and Redish, A.D. (2018). Deliberation and Procedural Automation on a Two-Step Task for Rats. Front. Integr. Neurosci. *12*.

Heien, M.L., Khan, A.S., Ariansen, J.L., Cheer, J.F., Phillips, P.E., Wassum, K.M., and Wightman, R.M. (2005). Real-time measurement of dopamine fluctuations after cocaine in the brain of behaving rats. Proc. Natl. Acad. Sci. U. S. A. *102*, 10023–10028.

Huotari, M., Gainetdinov, R., and Männistö, P.T. (1999). Microdialysis Studies on the Action of Tolcapone on Pharmacologically-Elevated Extracellular Dopamine Levels in Conscious Rats. Pharmacol. Toxicol. *85*, 233–238.

Huotari, M., Santha, M., Lucas, L.R., Karayiorgou, M., Gogos, J.A., and Männistö, P.T. (2002). Effect of Dopamine Uptake Inhibition on Brain Catecholamine Levels and Locomotion in Catechol-O-methyltransferase-Disrupted Mice. J. Pharmacol. Exp. Ther. *303*, 1309–1316.

Izenwasser, S., Werling, L.L., and Cox, B.M. (1990). Comparison of the effects of cocaine and other inhibitors of dopamine uptake in rat striatum, nucleus accumbens, olfactory tubercle, and medial prefrontal cortex. Brain Res. *520*, 303–309.

Käenmäki, M., Tammimäki, A., Myöhänen, T., Pakarinen, K., Amberg, C., Karayiorgou, M., Gogos, J.A., and Männistö, P.T. (2010). Quantitative role of COMT in dopamine clearance in the prefrontal cortex of freely moving mice: Quantitative role of COMT in the prefrontal cortex. J. Neurochem. *114*, 1745–1755.

Kaiser, R.H., Treadway, M.T., Wooten, D.W., Kumar, P., Goer, F., Murray, L., Beltzer, M., Pechtel, P., Whitton, A., Cohen, A.L., et al. (2018). Frontostriatal and Dopamine Markers of Individual Differences in Reinforcement Learning: A Multi-modal Investigation. Cereb. Cortex *28*, 4281–4290.

Karoum, F., Chrapusta, S.J., and Egan, M.F. (1994). 3-Methoxytyramine Is the Major Metabolite of Released Dopamine in the Rat Frontal Cortex: Reassessment of the Effects of Antipsychotics on the Dynamics of Dopamine Release and Metabolism in the Frontal Cortex, Nucleus Accumbens, and Striatum by a Simple Two Pool Model. J. Neurochem. *63*, 972–979.

Keithley, R.B., Mark Wightman, R., and Heien, M.L. (2009). Multivariate concentration determination using principal component regression with residual analysis. TrAC Trends Anal. Chem. *28*, 1127–1136.

Keithley, R.B., Carelli, R.M., and Wightman, R.M. (2010). Rank Estimation and the Multivariate Analysis of in Vivo Fast-Scan Cyclic Voltammetric Data. Anal. Chem. *82*, 5541–5551.

Kellendonk, C., Simpson, E.H., Polan, H.J., Malleret, G., Vronskaya, S., Winiger, V., Moore, H., and Kandel, E.R. (2006). Transient and Selective Overexpression of Dopamine D2 Receptors in

the Striatum Causes Persistent Abnormalities in Prefrontal Cortex Functioning. Neuron *49*, 603–615.

Koda, K., Ago, Y., Cong, Y., Kita, Y., Takuma, K., and Matsuda, T. (2010). Effects of acute and chronic administration of atomoxetine and methylphenidate on extracellular levels of noradrenaline, dopamine and serotonin in the prefrontal cortex and striatum of mice. J. Neurochem. *114*, 259–270.

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. *82*, 1–26.

Laatikainen, L.M., Sharp, T., Harrison, P.J., and Tunbridge, E.M. (2013). Sexually Dimorphic Effects of Catechol-O-Methyltransferase (COMT) Inhibition on Dopamine Metabolism in Multiple Brain Regions. PLoS ONE *8*, e61839.

Lammel, S., Hetzel, A., Häckel, O., Jones, I., Liss, B., and Roeper, J. (2008). Unique Properties of Mesoprefrontal Neurons within a Dual Mesocorticolimbic Dopamine System. Neuron *57*, 760–773.

Lammel, S., Ion, D.I., Roeper, J., and Malenka, R.C. (2011). Projection-Specific Modulation of Dopamine Neuron Synapses by Aversive and Rewarding Stimuli. Neuron *70*, 855–862.

Lammel, S., Lim, B.K., and Malenka, R.C. (2014). Reward and aversion in a heterogeneous midbrain dopamine system. Neuropharmacology *76*, 351–359.

Lapish, C.C., Ahn, S., Evangelista, L.M., So, K., Seamans, J.K., and Phillips, A.G. (2009). Tolcapone enhances food-evoked dopamine efflux and executive memory processes mediated by the rat prefrontal cortex. Psychopharmacology (Berl.) *202*, 521–530.

Lindvall, O., Björklund, A., and Divac, I. (1978). Organization of catecholamine neurons projecting to the frontal cortex in the rat. Brain Res. *142*, 1–24.

Männistö, P.T., and Kaakkola, S. (1999). Catechol-O-methyltransferase (COMT): Biochemistry, Molecular Biology, Pharmacology, and Clinical Efficacy of the New Selective COMT Inhibitors. Pharmacol. Rev. *51*, 593–628.

Matsumoto, M., Weickert, C.S., Akil, M., Lipska, B.K., Hyde, T.M., Herman, M.M., Kleinman, J.E., and Weinberger, D.R. (2003). Catechol O-methyltransferase mRNA expression in human and rat brain: evidence for a role in cortical neuronal function. Neuroscience *116*, 127–137.

Mazei, M.S., Pluto, C.P., Kirkbride, B., and Pehek, E.A. (2002). Effects of catecholamine uptake blockers in the caudate-putamen and subregions of the medial prefrontal cortex of the rat. Brain Res. *936*, 58–67.

Michael, D.J., and Wightman, R.M. (1999). Electrochemical monitoring of biogenic amine neurotransmission in real time. J. Pharm. Biomed. Anal. *19*, 33–46.

Miller, K.J., Botvinick, M.M., and Brody, C.D. (2017). Dorsal hippocampus contributes to model-based planning. Nat. Neurosci. *20*, 1269–1276.

Myöhänen, T.T., Schendzielorz, N., and Männistö, P.T. (2010). Distribution of catechol-O-methyltransferase (COMT) proteins and enzymatic activities in wild-type and soluble COMT deficient mice.

Nomikos, G.G., Damsma, G., Wenkstern, D., and Fibiger, H.C. (1990). In vivo characterization of locally applied dopamine uptake inhibitors by striatal microdialysis. Synapse *6*, 106–112.

Papageorgiou, G.K., Baudonnat, M., Cucca, F., and Walton, M.E. (2016). Mesolimbic Dopamine Encodes Prediction Errors in a State-Dependent Manner. Cell Rep. *15*, 221–228.

Peciña, S., Cagniard, B., Berridge, K.C., Aldridge, J.W., and Zhuang, X. (2003). Hyperdopaminergic mutant mice have higher “wanting” but not “liking” for sweet rewards. J. Neurosci. *23*, 9395–9402.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., and Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature *442*, 1042.

Pozzi, L., Invernizzi, R., Cervo, L., Vallebuona, F., and Samanin, R. (1994). Evidence that Extracellular Concentrations of Dopamine Are Regulated by Noradrenergic Neurons in the Frontal Cortex of Rats. J. Neurochem. *63*, 195–200.

Pycock, C.J., Kerwin, R.W., and Carter, C.J. (1980). Effect of lesion of cortical dopamine terminals on subcortical dopamine receptors in rats. Nature *286*, 74–77.

Raevskii, K.S., Gainetdinov, R.R., Budygin, E.A., Mannisto, P., and Wightman, M. (2002). Dopaminergic transmission in the rat striatum in vivo in conditions of pharmacological modulation. Neurosci. Behav. Physiol. *32*, 183–188.

Robinson, D.L., and Wightman, R.M. (2004). Nomifensine amplifies subsecond dopamine signals in the ventral striatum of freely-moving rats. J. Neurochem. *90*, 894–903.

Rothman, R.B., Mele, A., Reid, A.A., Akunne, H., Greig, N., Thurkauf, A., Rice, K.C., and Pert, A. (1989). Tight binding dopamine reuptake inhibitors as cocaine antagonists. FEBS Lett. *257*, 341–344.

Saddoris, M.P., Sugam, J.A., Stuber, G.D., Witten, I.B., Deisseroth, K., and Carelli, R.M. (2015). Mesolimbic Dopamine Dynamically Tracks, and Is Causally Linked to, Discrete Aspects of Value-Based Decision Making. Biol. Psychiatry *77*, 903–911.

Sesack, S.R., Hawrylak, V.A., Matus, C., Guido, M.A., and Levey, A.I. (1998). Dopamine axon varicosities in the prelimbic division of the rat prefrontal cortex exhibit sparse immunoreactivity for the dopamine transporter. J. Neurosci. *18*, 2697–2708.

Sharma, S., Hryhorczuk, C., and Fulton, S. (2012). Progressive-ratio Responding for Palatable High-fat and High-sugar Food in Mice. J. Vis. Exp.

Sharp, M.E., Foerde, K., Daw, N.D., and Shohamy, D. (2016). Dopamine selectively remediates 'model-based' reward learning: a computational approach. Brain *139*, 355–364.

Shnitko, T.A., and Robinson, D.L. (2014). Anatomical and pharmacological characterization of catecholamine transients in the medial prefrontal cortex evoked by ventral tegmental area stimulation. Synap. N. Y. N *68*, 131–143.

Sinkala, E., McCutcheon, J.E., Schuck, M.J., Schmidt, E., Roitman, M.F., and Eddington, D.T. (2012). Electrode calibration with a microfluidic flow cell for fast-scan cyclic voltammetry. Lab. Chip *12*, 2403.

Slifstein, M., Kolachana, B., Simpson, E.H., Tabares, P., Cheng, B., Duvall, M., Gordon Frankle, W., Weinberger, D.R., Laruelle, M., and Abi-Dargham, A. (2008). COMT genotype predicts cortical-limbic D1 receptor availability measured with [11C]NNC112 and PET. Mol. Psychiatry *13*, 821–827.

Slopsema, J.S., Van Der Gugten, J., and De Bruin, J.P.C. (1982). Regional concentrations of noradrenaline and dopamine in the frontal cortex of the rat: dopaminergic innervation of the prefrontal subareas and lateralization of prefrontal dopamine. Brain Res. *250*, 197–200.

Sulzer, D., Cragg, S.J., and Rice, M.E. (2016). Striatal dopamine neurotransmission: Regulation of release and uptake. Basal Ganglia *6*, 123–148.

Syed, E.C.J., Grima, L.L., Magill, P.J., Bogacz, R., Brown, P., and Walton, M.E. (2016). Action initiation shapes mesolimbic dopamine encoding of future rewards. Nat. Neurosci. *19*, 34–36.

Tammimaki, A., Aonurm-Helm, A., Kaenmaki, M., and Mannisto, P.T. (2016). Elimination of extracellular dopamine in the medial prefrontal cortex of conscious mice analysed using selective enzyme and uptake inhibitors. J. Physiol. Pharmacol. Off. J. Pol. Physiol. Soc. *67*, 301–309.

Tanda, G., Pontieri, F.E., Frau, R., and Di Chiara, G. (1997). Contribution of Blockade of the Noradrenaline Carrier to the Increase of Extracellular Dopamine in the Rat Prefrontal Cortex by Amphetamine and Cocaine. Eur. J. Neurosci. *9*, 2077–2085.

Tunbridge, E.M., Bannerman, D.M., Sharp, T., and Harrison, P.J. (2004). Catechol-O-Methyltransferase Inhibition Improves Set-Shifting Performance and Elevates Stimulated Dopamine Release in the Rat Prefrontal Cortex. J. Neurosci. *24*, 5331–5335.

Tunbridge, E.M., Harrison, P.J., and Weinberger, D.R. (2006). Catechol-o-Methyltransferase, Cognition, and Psychosis: Val158Met and Beyond. Biol. Psychiatry *60*, 141–151.

Tunbridge, E.M., Huber, A., M Farrell, S., Stumpenhorst, K., J Harrison, P., and E Walton, M. (2012). The role of catechol-O-methyltransferase in reward processing and addiction. CNS Neurol. Disord.-Drug Targets Former. Curr. Drug Targets-CNS Neurol. Disord. *11*, 306–323.

Valentini, V., Frau, R., and Di Chiara, G. (2004). Noradrenaline transporter blockers raise extracellular dopamine in medial prefrontal but not parietal and occipital cortex: differences with mianserin and clozapine. J. Neurochem. *88*, 917–927.

Walton, M.E., Behrens, T.E.J., Noonan, M.P., and Rushworth, M.F.S. (2011). Giving credit where credit is due: orbitofrontal cortex and valuation in an uncertain world. Ann. N. Y. Acad. Sci. *1239*, 14–24.

Weikop, P., Kehr, J., and Scheel-Krüger, J. (2007). Reciprocal effects of combined administration of serotonin, noradrenaline and dopamine reuptake inhibitors on serotonin and dopamine levels in the rat prefrontal cortex: the role of 5-HT1A receptors. J. Psychopharmacol. (Oxf.) *21*, 795–804.

Wunderlich, K., Smittenaar, P., and Dolan, R.J. (2012). Dopamine Enhances Model-Based over Model-Free Choice Behavior. Neuron *75*, 418–424.

Yavich, L., Forsberg, M.M., Karayiorgou, M., Gogos, J.A., and Mannisto, P.T. (2007). Site-Specific Role of Catechol-O-Methyltransferase in Dopamine Overflow within Prefrontal Cortex and Dorsal Striatum. J. Neurosci. *27*, 10196–10209.

Yin, H.H., Zhuang, X., and Balleine, B.W. (2006). Instrumental learning in hyperdopaminergic mice. Neurobiol. Learn. Mem. *85*, 283–288.

Yorgason, J.T., Jones, S.R., and España, R.A. (2011). Low and high affinity dopamine transporter inhibitors block dopamine uptake within 5 sec of intravenous injection. Neuroscience *182*, 125–132.

Young, J.W., and Geyer, M.A. (2010). Action of Modafinil—Increased Motivation Via the Dopamine Transporter Inhibition and D1 Receptors? Biol. Psychiatry *67*, 784–787.

Zhuang, X., Oosting, R.S., Jones, S.R., Gainetdinov, R.R., Miller, G.W., Caron, M.G., and Hen, R. (2001). Hyperactivity and impaired response habituation in hyperdopaminergic mice. Proc. Natl. Acad. Sci. *98*, 1982–1987.

Figure Legends

Figure 1: Effects of DAT blockade and COMT inhibition on the progressive ratio (PR) task

A within-subjects design was used to test drug effects; n=21. **A)** Depiction of the progression across trials of lever press ratios required to obtain reward during PR sessions. **B)** Schematic of experiment structure. Drug injections are indicated below the timeline. Timing of drug administration shown is for cohort 2 (see Methods). **C)** Number of responses on the active lever during the PR session under the four drug 1 & drug 2 treatment conditions: vehicle & vehicle (black), COMT inhibitor & vehicle (red), vehicle & DAT blocker (blue), and COMT inhibitor & DAT blocker (purple). Each data point shows the lever press session total for one animal. Boxplots show median and 25th and 75th percentiles; whiskers extend from the minimum to maximum value. Lever press data are shown on a $\log_{10}$ scale for clarity. **D)** As in (C), but for responses on the inactive lever. Data points again show lever press session totals for individual animals and are displayed on a $\log_{10}$ scale. **E)** As in (C), but for the latency to collect reward following its delivery. Each data point shows the cross-trial average latency for one animal. **F)** As in (C), but for the latency to re-engage with the task by recommencing lever pressing following the consumption of reward. Data points show cross-trial average latencies for individual animals.

Figure 2: DAT blockade impairs value learning during reward-guided decision making

A) Diagram of two-step task apparatus and trial events. **B)** Diagram of the task state space. **C)** Example session. Top panel: black line shows exponential moving average of subjects choices (tau = 8 trials). The correct choice (high or low) for each block is indicated by the grey bars. Middle panel: reward probabilities for each block for the left (red) and right (blue) sides. Bottom panel: transition probabilities linking first step actions to second step states, yellow line shows

probability of high→left and low→right transitions, purple line shows probability of high→right and low→left transitions. Colors in middle and bottom panel of (C) match those in (B). **D)** Trial rate as a function of session time. Trial rates were smoothed with a Gaussian of 5 minute standard deviation. Shaded areas show cross-subject SEM. **E)** Median second-step reaction times following common and rare transitions. The second-step reaction time is the time from choosing high or low to entering the active side port at the second step. Error bars show cross subject SEM. **F)** Choice probability trajectories around reversals in reward probabilities (left panel) and transition probabilities (right panel). Pale lines show cross-subject mean choice probability trajectory and dark lines show double exponential fit. **G)** Probability of repeating the first step choice on the next trial as a function of the trial outcome (rewarded or not) and experience state transition (common or rare).

Figure 2-1: Reinforcement learning modeling of baseline multi-step decision making behavior

Model comparison showing BIC score differences, relative to the best fitting model, for RL agents fitted to choices on baseline testing days (days with neither drug nor vehicle injections). The set of models considered were generated by adding or removing single features from the model found to best describe behavior on this task in Akam et al., 2017. Models are labeled on the x-axis by the feature that was added or removed.

Figure 3: COMT inhibition improves value learning during reward-guided decision making

A) Trial rate as a function of session time. Trial rates were smoothed with a Gaussian of 5 minute standard deviation. Shaded areas show cross-subject SEM. **B)** Median second-step reaction times following common and rare transitions. The second-step reaction time is the time from choosing high or low to entering the active side port at the second step. Error bars show

cross subject SEM. **C)** Choice probability trajectories around reversals in reward probabilities (left panel) and transition probabilities (right panel). Pale lines show cross-subject mean choice probability trajectory and dark lines show double exponential fit. **D)** Probability of repeating the first step choice on the next trial as a function of the trial outcome (rewarded or not) and experience state transition (common or rare).

Figure 4: Effects of COMT inhibition and DAT blockade on evoked dopamine release in the NAc

A) Location of recording and stimulating electrodes (see Figure 4-1). **B)** Illustration of features of dopamine transients quantified during analysis. **C)** Schematic of experiment structure. Drug injections are indicated below the timeline and time points of interest above it. **D)** An example recording made in the NAc core. The pseudocolor plot shows the recorded current as a function of the applied potential of the triangular waveform (y axis) over an 11sec period (x axis); timing and duration of stimulation indicated by thick black bar. The trace above the pseudocolor plot shows the extracted current attributable to dopamine release as a function of time. The cyclic voltammogram (current versus applied voltage; boxed inset) during the peak of the signal, 1.5sec after the time of stimulation, is consistent with dopamine release. **E)** Evoked dopamine release in the NAc in animals given the COMT inhibitor as drug 1 (red) compared to release in those given vehicle as drug 1 (black) 85min after the first injection. Release is normalized to the average pre-drug baseline peak height (equivalent to 100% on the y axis), binned over 15min centered on the time point of interest, and presented as mean ± SEM across animals within each drug group. Timing and duration of stimulation indicated by thick black bar. **F)** As in (E), but comparing release in animals given the DAT blocker as drug 2 (blue) with release in those given vehicle as drug 2 (black) 30min after the second injection. **G)** Quantification of the peak height of evoked dopamine release in NAc following administration of the COMT inhibitor. Left: peak height at the same time point shown in (E). Each animal's data is shown individually and is

normalized to its average pre-drug baseline peak height (equivalent to 100% on the y axis). Box plots show median and 25th and 75th percentiles; whiskers extend from the minimum to maximum value. Right: normalized peak height (mean ± SEM for each drug group) over the 90min following the first injection in animals that received the COMT inhibitor compared with those that received vehicle as drug 1. **H)** As in (G), but comparing data from animals that received the DAT blocker with data from those that received vehicle as drug 2; right-hand plot shows peak height over the 90min following the second injection. **I)** As in the left-hand plot of (G), but showing the quantification of the latency from stimulation to peak (left) and of the decay from the peak to T50 (right). Decay constant data are shown on a $\log_{10}$ scale for clarity. **J)** As in (I), but for the DAT blocker.

Figure 4-1: Electrode placements for NAc voltammetry recordings

A) Locations of electrolytic lesions made at recording electrode sites in the subset of animals in which lesions were made (11 out of 23 total animals). **B)** Locations of the ends of stimulating electrode tracts. **C)** Examples of a recording electrode lesion in the NAc (left) and a stimulating electrode tract terminating in the VTA (right), indicated by white arrowheads. The VTA histological section was stained for tyrosine hydroxylase and demonstrates the proximity of the stimulating electrode tract to dopaminergic cell bodies (black) in the midbrain. Numbers to the left of coronal sections in (A) and (B) indicate distance from bregma (mm). NAc: nucleus accumbens; PBP: parabrachial pigmented nucleus; PN: paranigral nucleus; SNc: substantia nigra pars compacta; SNr: substantia nigra pars reticulata; VTA: ventral tegmental area.

Figure 5: Effects of COMT inhibition and DAT blockade on evoked dopamine release in the MFC

A) Location of recording and stimulating electrodes (see Figure 5-1). **B)** An example recording made in the prelimbic MFC. The pseudocolor plot shows the recorded current as a function of

the applied potential of the triangular waveform (y axis) over a 6sec period (x axis); timing and duration of stimulation indicated by thick black bar. The trace above the pseudocolor plot shows the extracted current attributable to dopamine release as a function of time. The cyclic voltammogram (current versus applied voltage; boxed inset) during the peak of the signal, 1.4sec after the time of stimulation, is consistent with catecholamine release. **C)** Evoked dopamine release in the MFC in animals given the COMT inhibitor as drug 1 (red) compared to release in those given vehicle as drug 1 (black) 85min after the first injection. Release is normalized to the average pre-drug baseline peak height (equivalent to 100% on the y axis), binned over 15min centered on the time point of interest, and presented as mean ± SEM across animals within each drug group. Timing and duration of stimulation indicated by thick black bar. **D)** As in (C), but comparing release in animals given the DAT blocker as drug 2 (blue) with release in those given vehicle as drug 2 (black) 30min after the second injection. **E)** Quantification of the peak height of evoked dopamine release in MFC following administration of the COMT inhibitor. Left: peak height at the same time point shown in (C). Each animal's data is shown individually and is normalized to its average pre-drug baseline peak height (equivalent to 100% on the y axis). Box plots show median and 25th and 75th percentiles; whiskers extend from the minimum to maximum value. Right: normalized peak height (mean ± SEM for each drug group) over the 90min following the first injection in animals that received the COMT inhibitor compared with those that received vehicle as drug 1. **F)** As in (E), but comparing data from animals that received the DAT blocker with data from those that received vehicle as drug 2; right-hand plot shows peak height over the 90min following the second injection. **G)** As in the left-hand plot of (E), but showing the quantification of the latency from stimulation to peak (left) and of the decay from the peak to T50 (right). **H)** As in (G), but for the DAT blocker. **I)** The effect of amphetamine (magenta) on evoked dopamine release in MFC. Left: schematic of experiment structure. Middle: evoked dopamine release normalized to the average pre-drug baseline peak height, binned over 15min centered on the time point of interest (30min after drug injection), and

presented as mean ± SEM across animals within each group. Right: quantification of the signal decay, with each animal's data shown individually and normalized to its average pre-drug baseline decay constant.

Figure 5-1: Electrode placements for MFC voltammetry recordings

A) Locations of electrolytic lesions made at recording electrode sites in the subset of animals in which lesions were made (16 out of 34 total animals). **B)** Locations of the ends of stimulating electrode tracts. **C)** Examples of a recording electrode lesion in the MFC (left) and a stimulating electrode tract terminating in the VTA (right), indicated by white arrowheads. The VTA histological section was stained for tyrosine hydroxylase and demonstrates the proximity of the stimulating electrode tract to dopaminergic cell bodies (black) in the midbrain. Numbers to the left of coronal sections in (A) and (B) indicate distance from bregma (mm). Cgl: cingulated cortex; MO: medial orbital cortex; PBP: parabrachial pigmented nucleus; PN: paranigral nucleus; PrL: prelimbic cortex; SNc: substantia nigra pars compacta; SNr: substantia nigra pars reticulata; VTA: ventral tegmental area.

Figure 5-2: Effects of NET blockade on evoked dopamine release in the MFC

A) The effect of the NET blocker atomoxetine (green) on evoked dopamine release in MFC. Left: schematic of experiment structure. Middle: evoked dopamine release normalized to the average pre-drug baseline peak height, binned over 15min centered on the time point of interest (60min after drug injection), and presented as mean ± SEM across animals within each group. Right: quantification of the signal decay, with each animal's data shown individually and normalized to its average pre-drug baseline decay constant.

Tables

Table 1: Multi-step decision making task parameter changes over training.

Session number	Reward size (µl)	Transition probabilities (common / rare)	Reward probabilities (good / bad side)
1	4 or 8	0.9 / 0.1	First 40 trials all rewarded, subsequently 0.9 / 0.1
2 - 18	4 or 8	0.9 / 0.1	0.9 / 0.1
19 - 21	8	0.8 / 0.2	0.9 / 0.1
22 - 23	6.5	0.8 / 0.2	0.9 / 0.1
24 - 26	6.5	0.8 / 0.2	0.8 / 0.2
27+	4	0.8 / 0.2	0.8 / 0.2

Table 2: Voltammetry experiment group sizes.

Region & drug group	Group size for each drug treatment	
NAc COMT inhibition	Vehicle: n = 12 (n = 5 Veh/Veh, n = 7 Veh/GBR)	Tolcapone: n = 11 (n = 6 Tolc/Veh, n = 5 Tolc/GBR)
NAc DAT blockade	Vehicle: n = 11 (n = 5 Veh/Veh, n = 6 Tolc/Veh)	GBR-12909: n = 12 (n = 7 Veh/GBR, n = 5 TolcGBR
MFC COMT inhibition	Vehicle: n = 11 (n = 6 Veh/Veh, n = 5 Veh/GBR)	Tolcapone: n = 14 (n = 6 Tolc/Veh, n = 8 Tolc/GBR)
MFC DAT blockade	Vehicle: n = 12 (n = 6 Veh/Veh, n = 6 Tolc/Veh)	GBR-12909: n = 13 (n = 5 Veh/GBR, n = 8 Tolc/GBR)
MFC amphetamine	Amphetamine: n = 7 (n = 1 Veh/Veh/Amph, n = 2 Tolc/Veh/Amph, n = 4 Amph only)	

Table 3: DAT blockade reversal analysis fit for the multi-step decision making task. Parameters of double exponential fit to reversals in reward and transition probabilities under GBR-12909 and vehicle.

Reversal type	**Condition**	**Parameter**		
		tau_F	tau_S	w_F
Reward	*Vehicle*	2.58	280.12	0.70
	GBR-12909	15.68	7842.5	0.89
	p value	0.0012	0.0008	0.98
Transition	*Vehicle*	0.83	35.95	0.44
	GBR-12909	3.58	62.27	0.64
	p value	0.23	0.69	0.44

Table 4: COMT inhibition reversal analysis fit for the multi-step decision making task. Parameters of double exponential fit to reversals in reward and transition probabilities under tolcapone and vehicle.

Reversal type	**Condition**	**Parameter**		
		tau_F	tau_S	w_F
Reward	*Vehicle*	13.6	54.24	0.57
	Tolcapone	3.98	39.27	0.48
	p value	0.0496	0.84	0.84
Transition	*Vehicle*	5.03	2514	0.90
	Tolcapone	4.42	2207	0.76
	p value	0.72	0.75	0.3

Figure 1:

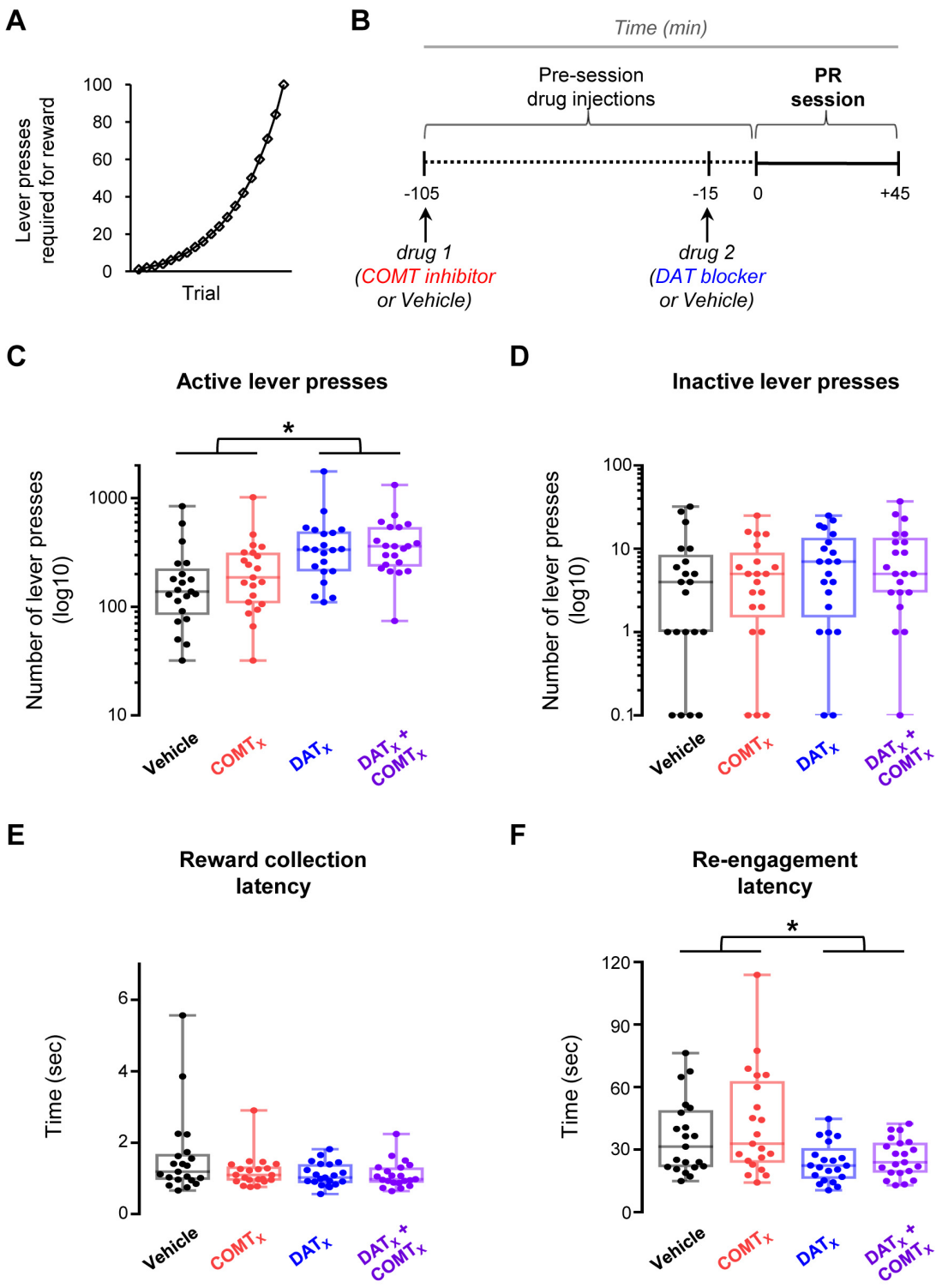
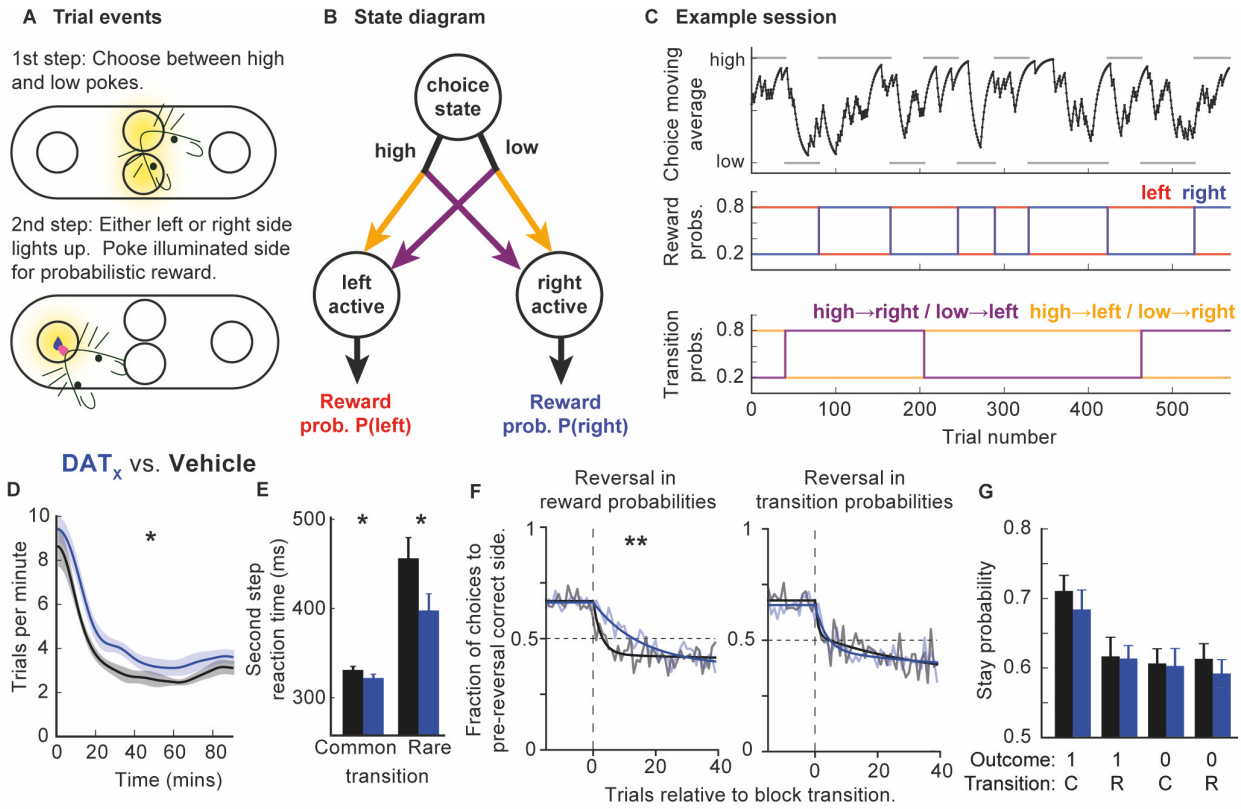

A
Lever presses required for reward
100
80
60
40
20
0
Trial
B
Time (min)
Pre-session drug injections
PR session
-105
-15
0
+45
drug 1 (COMT inhibitor or Vehicle)
drug 2 (DAT blocker or Vehicle)
C
Active lever presses
Number of lever presses (log10)
1000
100
10
*
Vehicle
COMTx
DATx
DATx + COMTx
D
Inactive lever presses
Number of lever presses (log10)
100
10
1
0.1
Vehicle
COMTx
DATx
DATx + COMTx
E
Reward collection latency
Time (sec)
6
4
2
0
Vehicle
COMTx
DATx
DATx + COMTx
F
Re-engagement latency
Time (sec)
120
90
60
30
0
*
Vehicle
COMTx
DATx
DATx + COMTx

Figure 2:


A Trial events
1st step: Choose between high and low pokes.
2nd step: Either left or right side lights up. Poke illuminated side for probabilistic reward.
B State diagram
choice state
high
low
left active
right active
Reward prob. P(left)
Reward prob. P(right)
C Example session
Choice moving average
high
low
Reward probs.
left right
0.8
0.2
Transition probs.
high→right / low→left
high→left / low→right
0.8
0.2
0
100
200
300
400
500
Trial number
DATx vs. Vehicle
D
Trials per minute
10
8
6
4
2
0
0
20
40
60
80
Time (mins)
*
E
Second step reaction time (ms)
500
400
300
*
*
Common
Rare
transition
F
Reversal in reward probabilities
Reversal in transition probabilities
Fraction of choices to pre-reversal correct side.
1
0.5
0
**
0
20
40
Trials relative to block transition.
G
Stay probability
0.8
0.7
0.6
0.5
Outcome: 1 1 0 0
Transition: C R C R

Extended figure 2-1:

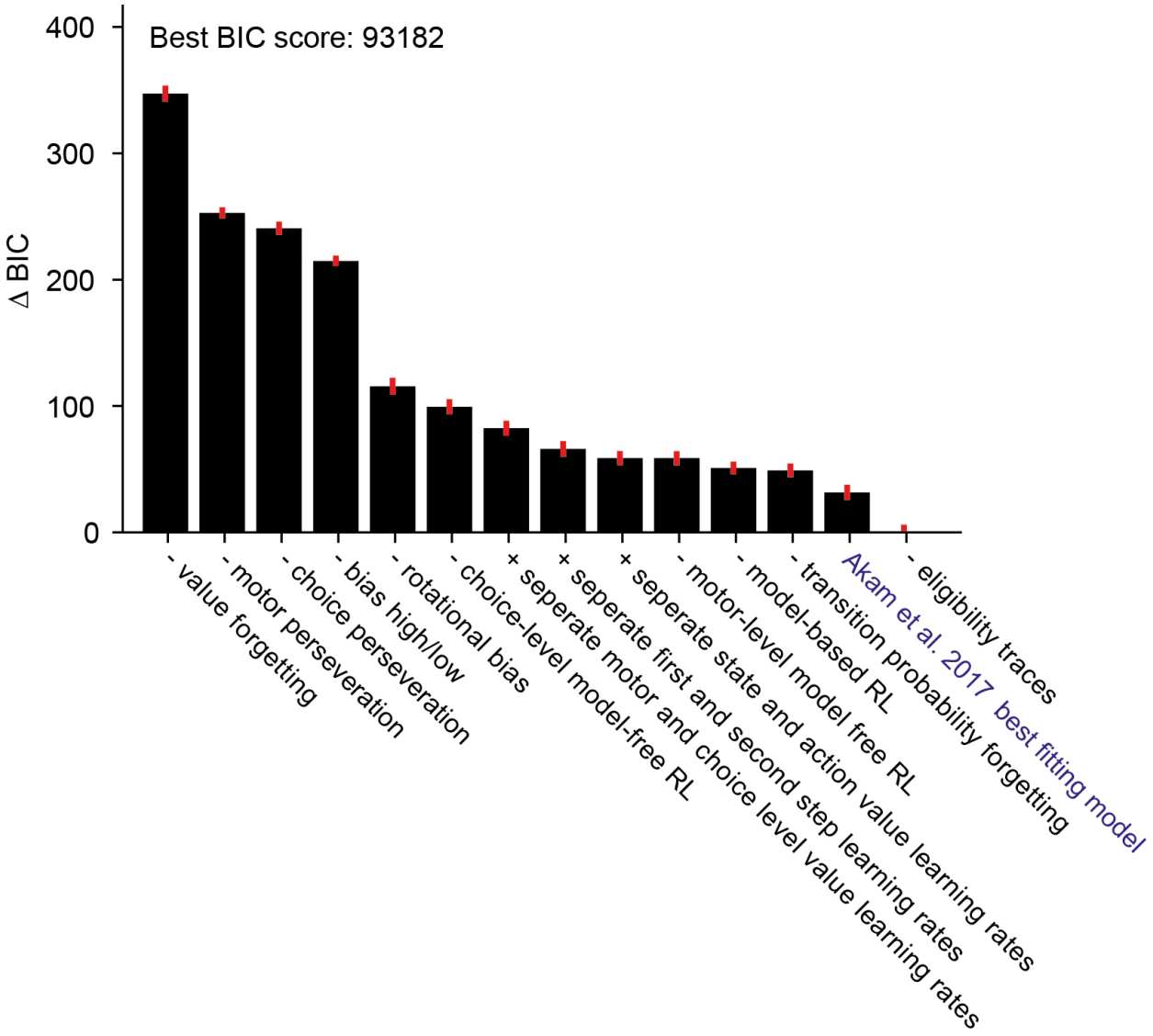


Figure 3:

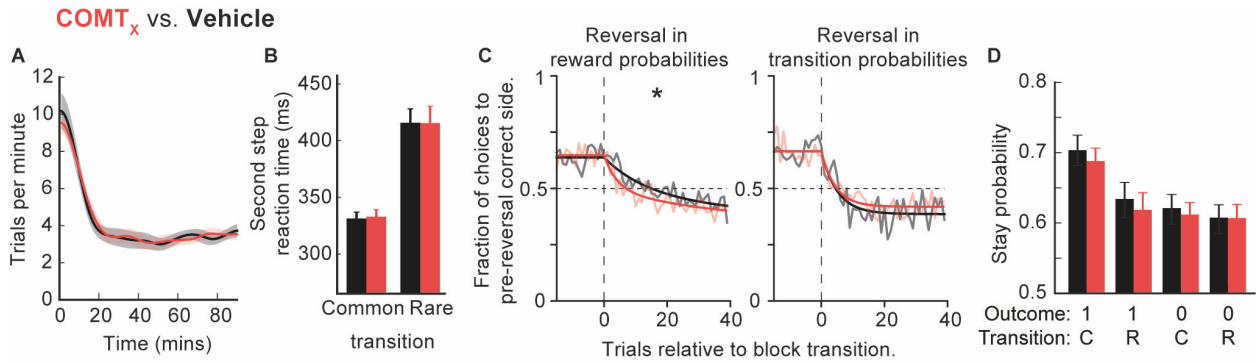
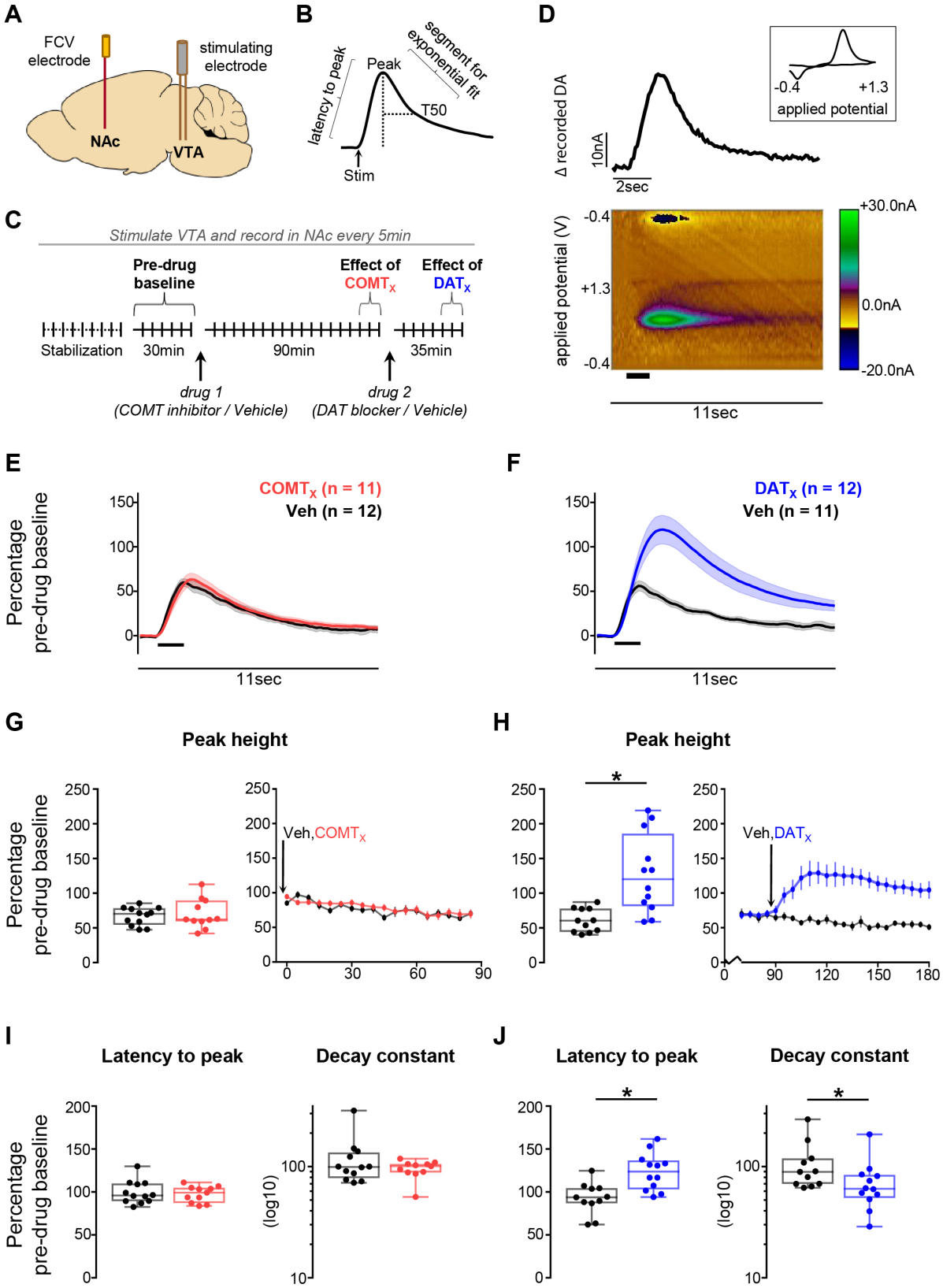

COMTx vs. Vehicle
A
Trials per minute
Time (mins)
B
Second step reaction time (ms)
Common Rare
transition
C
Reversal in reward probabilities
Reversal in transition probabilities
Fraction of choices to pre-reversal correct side.
Trials relative to block transition.
D
Stay probability
Outcome: 1 1 0 0
Transition: C R C R

Figure 4:


A
FCV electrode
stimulating electrode
NAc
VTA
B
latency to peak
Peak
segment for exponential fit
T50
Stim
D
Δ recorded DA
10nA
2sec
-0.4
+1.3
applied potential
applied potential (V)
+30.0nA
0.0nA
-20.0nA
11sec
C
Stimulate VTA and record in NAc every 5min
Pre-drug baseline
Effect of COMTx
Effect of DATx
Stabilization
30min
90min
35min
drug 1 (COMT inhibitor / Vehicle)
drug 2 (DAT blocker / Vehicle)
E
COMTx (n = 11)
Veh (n = 12)
Percentage pre-drug baseline
11sec
F
DATx (n = 12)
Veh (n = 11)
11sec
G
Peak height
Veh,COMTx
H
Peak height
Veh,DATx
I
Latency to peak
Decay constant
(log10)
J
Latency to peak
Decay constant

Extended figure 4-1:

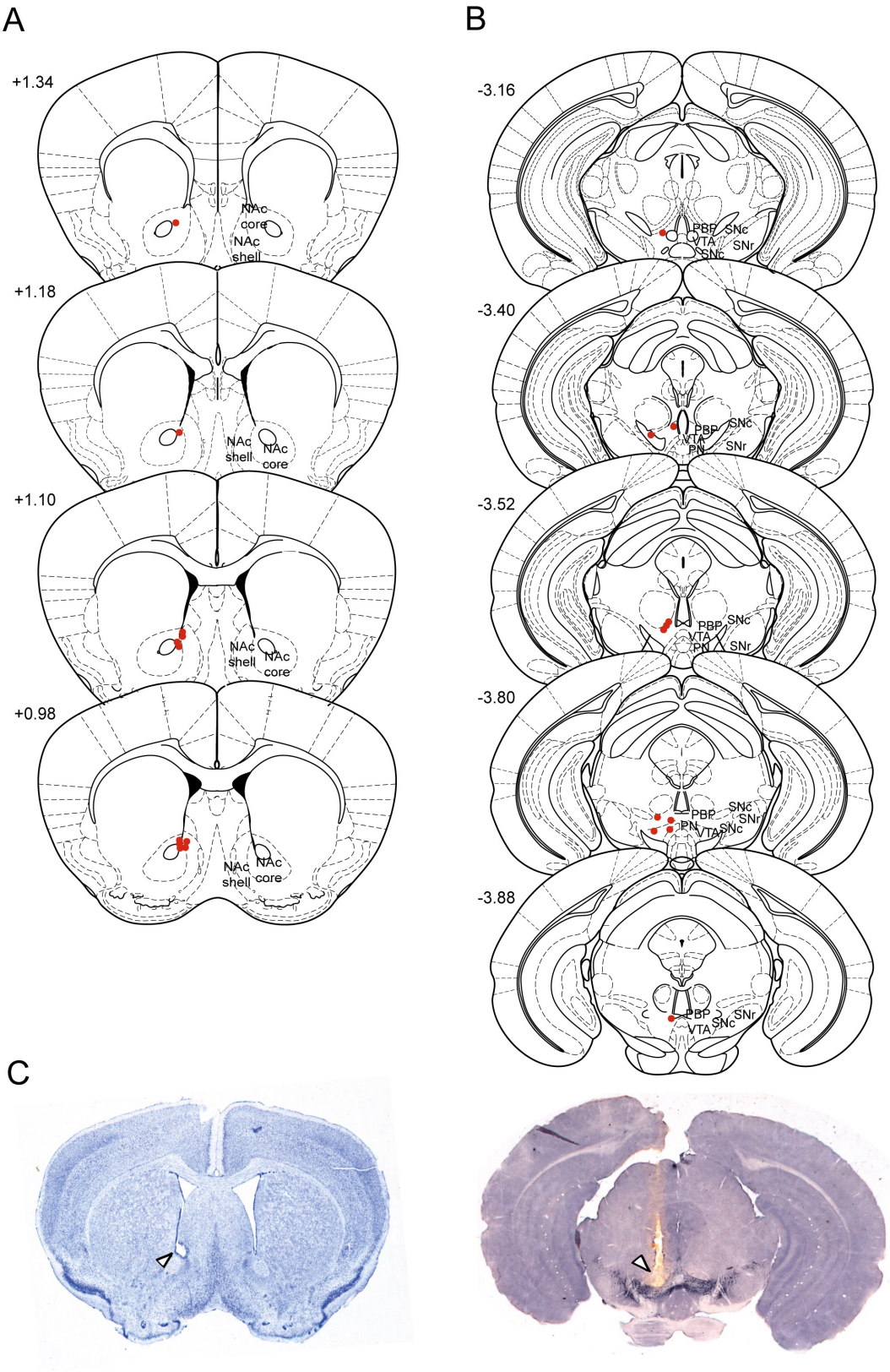
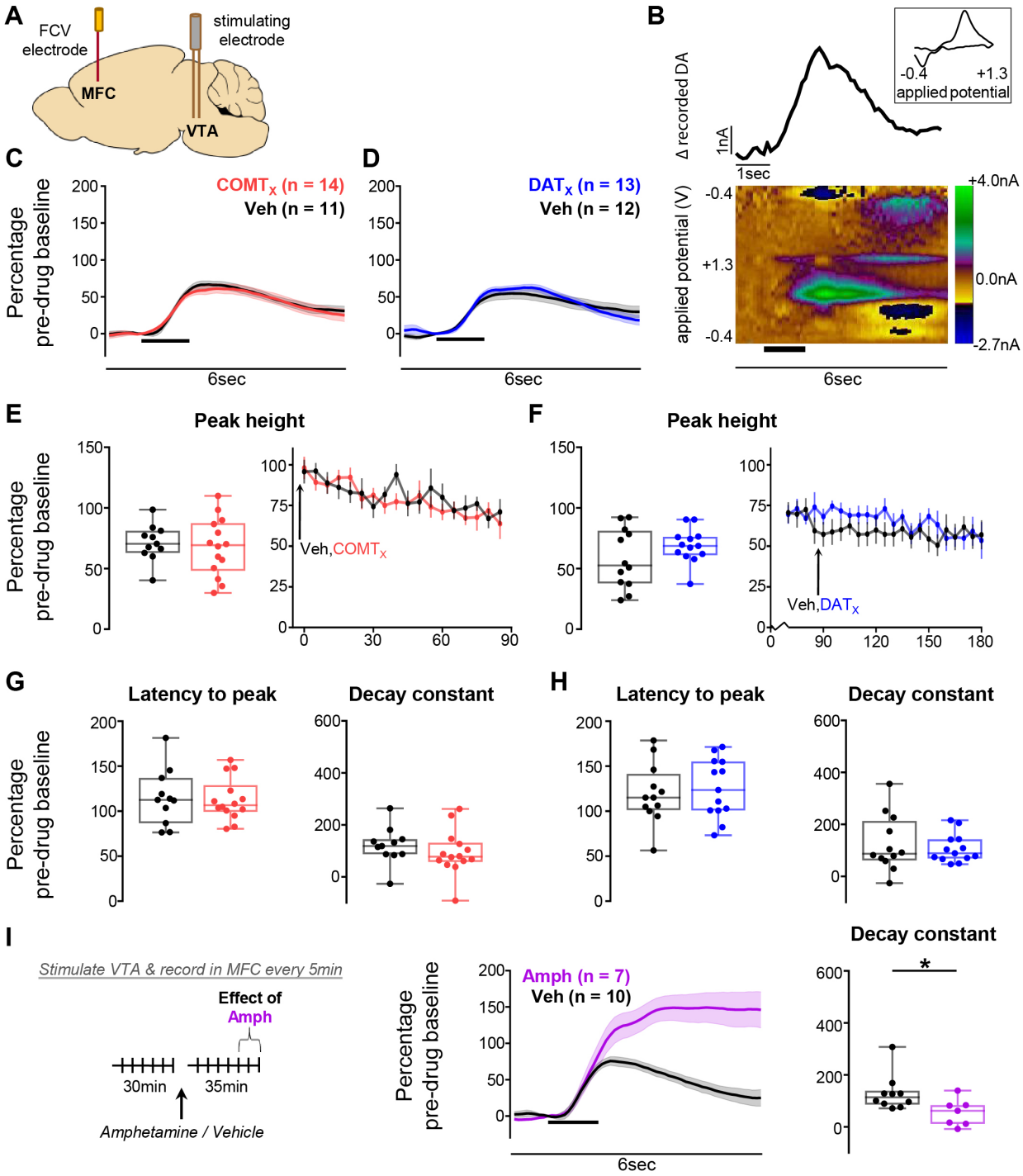
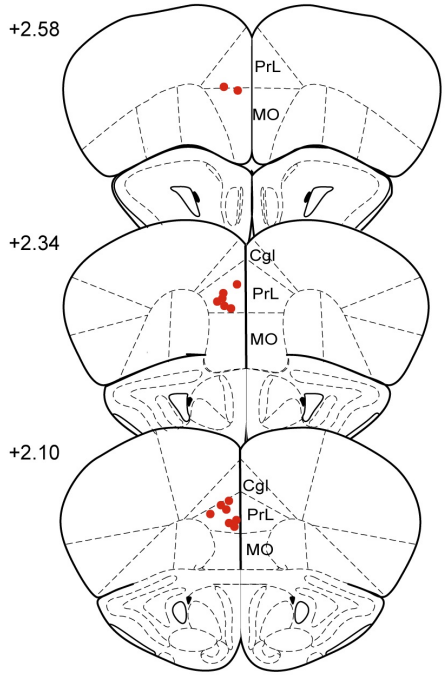


Figure 5:

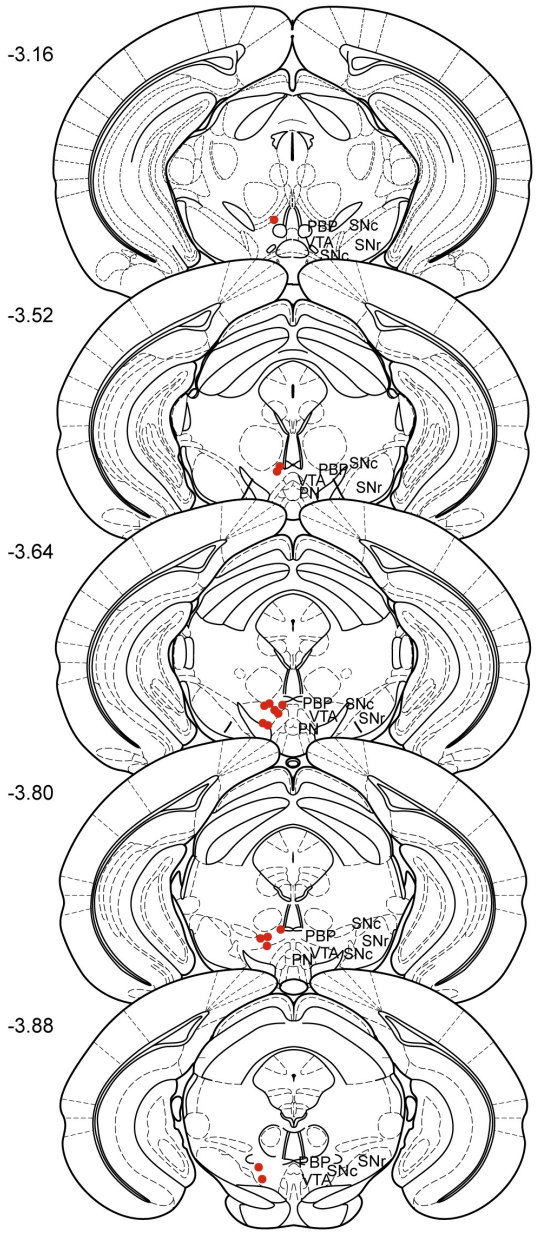

A
FCV electrode
stimulating electrode
MFC
VTA
B
Δ recorded DA
1nA
1sec
-0.4
+1.3
applied potential
applied potential (V)
+4.0nA
0.0nA
-2.7nA
6sec
C
Percentage pre-drug baseline
COMTX (n = 14)
Veh (n = 11)
D
DATX (n = 13)
Veh (n = 12)
E
Peak height
Veh,COMTX
F
Peak height
Veh,DATX
G
Latency to peak
Decay constant
H
Latency to peak
Decay constant
I
Stimulate VTA & record in MFC every 5min
Effect of Amph
30min
35min
Amphetamine / Vehicle
Amph (n = 7)
Veh (n = 10)
Decay constant
*

Extended figure 5-1:

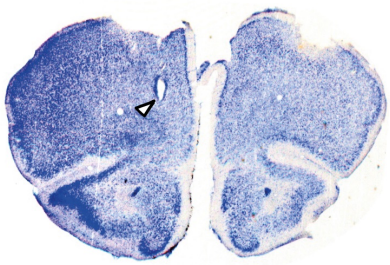
A

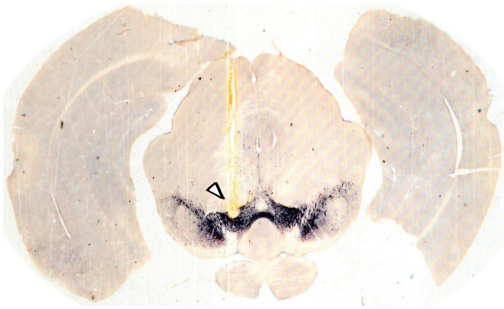
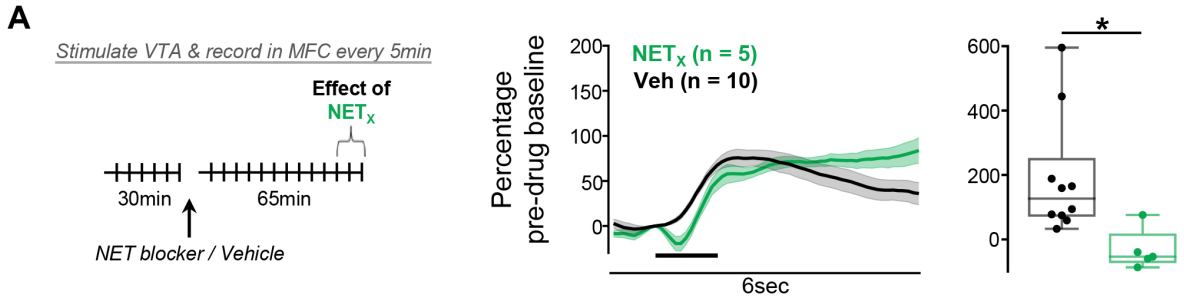


B



C





Extended figure 5-2: