Published in final edited form as:
Neurobiol Learn Mem. 2015 November ; 125: 135–145. doi:10.1016/j.nlm.2015.08.011.

Phasic dopamine release induced by positive feedback predicts individual differences in reversal learning

Marianne Klanker[1,2], **Tessa Sandberg**[1], **Ruud Joosten**[1], **Ingo Willuhn**[1,2], **Matthijs Feenstra**[1,2], and **Damiaan Denys**[1,2]

[1]Netherlands Institute for Neuroscience, Institute of the Royal Netherlands Academy of Arts and Sciences, Meibergdreef 47, 1105 BA Amsterdam, The Netherlands [2]Department of Psychiatry, Academic Medical Center, University of Amsterdam, Postbus 22660, 1100 DD Amsterdam, The Netherlands

Abstract

Striatal dopamine (DA) is central to reward-based learning. Less is known about the contribution of DA to the ability to adapt previously learned behavior in response to changes in the environment, such as a reversal of response-reward contingencies. We hypothesized that DA is involved in the rapid updating of response-reward information essential for successful reversal learning. We trained rats to discriminate between two levers, where lever availability was signaled by a non-discriminative cue. Pressing one lever was always rewarded, whereas the other lever was never rewarded. After reaching stable discrimination performance, a reversal was presented, so that the previously non-rewarded lever was now rewarded and vice versa. We used fast-scan cyclic voltammetry to monitor DA release in the ventromedial striatum. During discrimination performance (pre-reversal), cue presentation induced phasic DA release, whereas reward delivery did not. The opposite pattern was observed post-reversal: Striatal DA release emerged after reward delivery, while cue-induced release diminished. Trial-by-trial analysis showed rapid reinstatement of cue-induced DA release on trials immediately following initial correct responses. This effect of positive feedback was observed in animals that learned the reversal, but not in ‘non-learners’. In contrast, neither pre-reversal responding and DA signaling, nor post-reversal DA signaling in response to negative feedback differed between learners and non-learners. Together, we show that phasic DA dynamics in the ventromedial striatum encoding reward-predicting cues are associated with positive, during reversal learning. Furthermore, these signals predict individual differences in learning that are not present prior to reversal, suggesting a distinct role for dopamine in the adaptation of previously learned behavior.

Keywords

Dopamine; reversal learning; fast-scan cyclic voltammetry; feedback learning; cognitive flexibility

Contact Information: Marianne Klanker, Netherlands Institute for Neuroscience (NIN), Meibergdreef 47, 1105 BA Amsterdam, The Netherlands, Phone: +31 20 5665489, Fax: +31 20 5666121, m.klanker@nin.knaw.nl; marianneklanker@yahoo.com.

1 Introduction

Throughout the day, we perform numerous behavioral actions to pursue things we desire or to prevent adverse events. In a constantly changing environment, this behavior has to be adaptive and flexible. One form of adaptive behavior is reversal learning, which requires the ability to use negative feedback to inhibit a learned response that was previously rewarded and at the same time to use positive feedback to switch to a response that was previously unrewarded. The ability to adapt goal-directed behavior to changes in the environment depends on frontal and striatal regions, as demonstrated in rodents (Castane et al. 2010; McAlonan and Brown 2003), non-human primates (Dias et al. 1996; Clarke et al. 2008) and humans (Bellebaum et al. 2008). Indeed, compromised frontostriatal integrity results in deficient adaptive behavior and cognitive dysfunctions in various neurological and psychiatric disorders, such as obsessive-compulsive disorder (OCD), drug addiction and Parkinson's disease (Chamberlain et al. 2006; Cools et al. 2001; Ceaser et al. 2008; Yerys et al. 2009; Verdejo-Garcia et al. 2006; Millan et al. 2012).

Dopamine (DA) is an important neuromodulator in frontostriatal networks. Striatal DA release facilitates reward learning and mediates approach behavior towards rewards. Specifically, DA neurons increase firing in response to unexpected rewards and reward-predicting stimuli and decrease firing when an expected reward is omitted (Schultz et al. 1997; Pan et al. 2005). These and other findings suggest that DA neurons encode a 'reward-prediction error' that serves as a teaching signal to guide behavior (Montague et al. 1996; Schultz et al. 1997; Waelti et al. 2001; Steinberg et al. 2013). It is known that burst firing of DA neurons facilitates initial learning of response-reward associations (Zweifel et al. 2009). Consistently, selective optogenetic activation of DA neurons affects operant responding in a similar manner as positive feedback does (Kim et al. 2012; Witten et al. 2011), and pharmacological and genetic manipulations demonstrate DA-mediated regulation of adaptive behavior (Clarke et al. 2011; Haluk and Floresco 2009; Laughlin et al. 2011; Klanker et al. 2013). DA may not only play such role during initial learning, but also in adaptation of established operant responding. Indeed, optogenetic activation of DA neurons (simulating positive feedback) can also mediate reversal of reward-seeking behavior (Adamantidis et al. 2011). However, theories of reinforcement learning and models of DA-mediated prediction errors suggest that fluctuations in striatal DA concentration mediate the effects of both positive and negative feedback on learning (Hong and Hikosaka 2011; Frank and Claus 2006). It remains to be shown whether DA mediates the effects of negative feedback during behavioral adaption following an unexpected switch in reward contingencies and whether striatal DA reflects the receipt of feedback on a trial-by-trial basis. Therefore, we used fast-scan cyclic voltammetry (FSCV) in behaving animals (Millar et al. 1985) to monitor rapid changes in DA release in the ventromedial striatum during an operant spatial reversal learning task, in which rats had to adapt goal-directed behavior following a reversal of response-reward contingencies. Our results show that phasic cue-evoked DA signals were promptly updated following positive, but not negative feedback. Furthermore, we observed individual differences in the extent to which the receipt of positive feedback updated the reward-predicting DA signal on subsequent trials, where DA signalling predicted successful reversal learning. Thus, our findings suggest that phasic DA in the ventromedial striatum

provides a positive feedback signal to facilitate adaptation of previously learned behavior following a behavioral switch.

2 Materials and Methods

2.1 Animals

Male Wistar rats (Charles River) were housed in a controlled environment under a reversed day-night schedule (white lights: 7 p.m.-7 a.m.). Rats were food-restricted (16 grams animal/day), with unlimited access to water during behavioral training. All experiments were approved by the Animal Experimentation Committee of the Royal Netherlands Academy of Arts and Sciences and were carried out in agreement with Dutch laws (Wet op de Dierproeven, 1996) and European regulations (Guideline 86/609/EEC).

2.2 Surgery

Rats (weighing ~300 grams) were anesthetized with Isoflurane (induction: 3%, maintenance: 1.8-2.5 %) and placed in a stereotactic frame. Subcutaneous Metacam® (Meloxicam, 1 mg/kg, Boehringer-Ingelheim, Germany) was given as analgesic. A guide cannula (custom made, NIN mechanical workshop) was implanted above the nucleus accumbens (anterioposterior +1.3 mm, lateral ±1.3 mm relative to bregma (Paxinos and Watson 2007). An Ag/AgCl reference electrode was implanted in the contralateral hemisphere. A bipolar stimulation electrode (Plastics one, Roanoke, VA, USA) was lowered into the ventral tegmental area (AP -5.2mm, ML -1.0 mm, DV – 8-9 mm from skull). The guide cannula, reference electrode and stimulating electrode were fixed to the skull with screws and dental cement. A removable stylet was used to close the cannula after surgery. Unlimited access to food and water was provided the day before surgery and during post-operative recovery. Rats were individually housed following surgery.

2.3 Behavioral training

All behavioral testing (see Table 1 for training phases) was performed in a custom made operant chamber (40 x 40 x 40 cm, NIN mechanical workshop) with MED Associates parts (Med Associates, Sandown Scientific, Hampton, UK). Two retractable levers were placed left and right from a food dispenser. Cue lights were positioned above both levers. Nose-pokes in the food dispenser to retrieve sucrose pellets (Dustless precision pellets®, 45 mg, Bio-Serv) were detected by an infrared sensor. During shaping sessions, rats were trained to press a lever for a food reward. Rats were randomly presented with the right or left cue light and corresponding lever (extended after a 2 sec delay). After a lever press, the lever was retracted, the cue light switched off and a sucrose pellet was delivered in the food dispenser. In case of an omission, cue and lever presentation ended after 30 seconds. Shaping sessions consisted of 32 trials, with a 10 or 20 second interval. Rats received up to three shaping sessions per day with an inter-session interval of 2-3 hours. After reaching a 90% correct response criterion (measured over the complete session), rats continued in a 64-trial shaping session. Rats then progressed to spatial discrimination learning (120 trials per session, variable inter-trial intervals (15/25/35/45 seconds)). On every trial both cue lights were illuminated and two seconds later both levers were extended into the operant chamber. Thus, cue lights did not signal which side was rewarded, but indicated that reward was available

provided the correct choice was made. Responding to the lever on one side was rewarded, while responding to the other lever on the other side was never rewarded. The rewarded side was counterbalanced between rats. If rats did not make a lever-press within 10 seconds, levers were retracted and the trial was scored as omission. Rats received one session per day. If rats did not reach a 90% correct response criterion during the second discrimination session, a third, shorter discrimination session (max 64 trials) was presented to allow them to get to the 90% response criterion without overtraining them. Reversal sessions consisted of 120 trials, with variable inter-trial intervals (15/25/35/45 seconds). The reversal was presented randomly between the 16th and 32nd trial in the session, so that a response to the previously non-rewarded lever was now rewarded and vice versa. The reversal was not cued to the animals; instead animals had to use the change in feedback to adapt responding.

2.4 Fast-scan cyclic voltammetry

FSCV was used to record DA changes during discrimination and reversal learning. For FSCV recordings (Phillips et al. 2003), a micromanipulator holding a glass-enclosed carbon fiber (electrode tip: diameter 7μm, length 100-150 μm) and attached to a head-mounted amplifier was used to record DA concentration in the ventromedial striatum. As a clear distinction between nucleus accumbens core and the ventromedial part of the caudate putamen cannot be made based on structural, functional, or connectivity differences (Zahm and Brog 1992; Voorn et al. 2004), we recorded from both regions. A potential of -0.4 V was applied to the carbon fiber electrode (vs an Ag/AgCl reference electrode). Resting potential changed to 1.3 V and back to resting potential in a triangular waveform (8.5 msec) every 100 msec. Redox reactions of DA molecules in vicinity of the electrode at specific applied potentials (0.6V oxidation, -0.2 V reduction) in the waveform result in the release of electrons that can be measured as current at the carbon fiber electrode. A fresh carbon fiber electrode was inserted before each recording session. Placement of the electrode in a DA-rich region was verified by presence of spontaneous DA release events (‘transients’ (Wightman and Robinson 2002)) and/or a time-locked DA response to the presentation of unexpected food pellets. After the behavioral session, VTA Stimulation (8, 12, 16 pulses, 30 Hz frequency, 125 μA intensity, 4 msec pulse width) was performed to construct a training set for chemometric analysis from electrically stimulated DA release and pH changes. Chemometric analysis was used to identify DA from other electro-active species (Heien et al. 2004; Keithley et al. 2009). Use of acutely inserted glass electrodes before each recording sessions limited the amount of recordings possible in the same animals. Therefore, most animals only contributed voltammetry data to either one of the two discrimination sessions or the reversal learning session making it impossible to correlate changes in DA release during discrimination sessions with DA release of the same animals during reversal learning. After the last behavioral session, a stainless steel lesion electrode was inserted to the recording location and a 100 μA direct current was passed through the electrode to mark the final placement of the electrode. Rats were deeply anesthetized using an overdose of pentobarbital (Erasmus MC pharmacy, Rotterdam, the Netherlands) and decapitated. Brains were removed and frozen. For histological verification of electrode placement, 40 μm coronal slices were cut on a cryostat and stained with cresyl violet.

2.5 Data analysis and statistics

The following behavioral measures were analyzed: number/percentage of responses to the rewarded and non-rewarded lever and latency to lever press. For the reversal session, a cumulative response record was plotted to visualize learning over the course of the session. Based on performance in the reversal session rats were divided into groups of 'learners' (>10 correct responses after presentation of reversal) and 'non-learners' (<10 correct responses after presentation of reversal). For the reversal session, a change point in behavior was defined as the point where the learning curve (cumulative number of correct responses) deviates maximally from a straight line drawn from the origin to the last point in the cumulative line (Gallistel et al, 2004). One change point was defined for each learning curve. Behavioral data was analyzed with independent t-tests (discrimination learning) or repeated measures ANOVA (reversal session; reversal stage (before/after) as repeated measure).

For analysis of the DA recordings, trials were averaged over a behavioral session for each rat, then averaged over rats. For cue-evoked responses, baseline value was an average of the measurements during 5 seconds before cue onset. Peak values were the maximal value in a 2 second window following cue presentation. For DA responses to reward delivery, DA following 4 seconds after lever press was averaged (for reward delivery, an average measure rather than peak value was chosen as less data points were available in period following lever press due to noise). For a more detailed analysis of DA changes related to reward delivery after reversal presentation, we took one-sec bins of the first four seconds following lever press for trials after reversal presentation. Here, average DA response in consecutive one-sec bins was compared to average DA during baseline period (first four seconds in trial, to exclude DA changes following cue presentation). Repeated measures ANOVAs were used to analyze changes in DA release during discrimination and reversal learning. Bonferroni-corrections were applied for post-hoc t-tests when appropriate. When assumption of sphericity was violated, Greenhouse-Geisser or Huyn-Feldt corrections were applied were appropriate. Independent t-tests were used to compare group differences (Welch t statistic reported when assumption of homogeneity of variances was violated). Statistical analyses were performed with SPSS Statistics 21 (IBM, Armonk, NY). Statistical significance was set to $p<0.05$.

3 Results

3.1 Histology

Fig. 1 illustrates the recording sites in the ventromedial striatum (dorsomedial accumbens core and ventromedial regions caudate putamen in right hemisphere). Final group sizes for animals included in FSCV data: first discrimination session n=11, second discrimination n=8, reversal session n=21. A statistical comparison of DA responses at the different locations used in the reversal session revealed no significant differences.

3.2 Phasic DA changes in the ventromedial striatum during spatial discrimination learning

After learning to lever-press for reward in several shaping sessions, rats were trained on a spatial discrimination paradigm: a lever-press on one lever was always rewarded, whereas a lever-press on the other lever was never rewarded. Number of rewarded responses increased from first (D1) to second (D2) discrimination session ($t(17)=-5.228$, $p<0.001$), whereas the number of non-rewarded responses decreased ($t(17)=6.089$, $p<0.001$). Response times did not differ between rewarded or non-rewarded responses. DA increased relative to baseline at time of cue onset during both discrimination sessions (D1: $F(1,20)=71.907$, $p<0.001$; D2: $F(1,14)=46.355$, $p<0.001$), but was not different between rewarded and non-rewarded trials, therefore did not predict whether animals made a correct or incorrect response. Thus, in our task cue-evoked DA did not encode the chosen response, but signaled the availability of reward (Roesch et al. 2007; Sugam et al. 2012). Cue-evoked DA was not different between trials that followed a correct response and trials following an incorrect response. No significant DA response to reward delivery was found during discrimination sessions. During shaping sessions, prior to discrimination learning, rats were trained to press a lever for reward. As (repeatedly) shown by others (Wassum et al. 2012; Roitman et al. 2004), the shift of DA release from time of reward delivery to time of reward-predicting cue likely already takes place during acquisition of instrumental behavior, prior to discrimination learning.

3.3 Dynamic changes in DA release during reversal learning

During the reversal session, the rewarded lever was switched at a random time point in the session. Rats then had to use feedback (i.e. previously rewarded response no longer rewarded; previously non-rewarded response now rewarded) to adapt their responding because the reversal was not cued. Before reversal, rats had a clear preference to press the rewarded lever. After reversal presentation, the number of rewarded responses decreased, whereas non-rewarded responses increased (reversal*reward interaction ($F(1,40)=299.238$, $p<0.001$; Fig. 2A). In addition, response latencies were significantly longer after reversal than before ($F(1,33)=7.836$, $p=0.008$), but did not differ between rewarded and non-rewarded responses. Fig. 2A shows the gradual adaptation of response behavior during reversal learning (session divided into blocks of 8 trials). Rewarded responses gradually increased (main effect for block ($F(3.527,70.545)=7.979$, $p<0.001$, simple contrasts vs block 1: differences from block 7 onwards), whereas non-rewarded responses gradually decreased ($F(10,200)=18.145$, $p<0.001$, simple contrasts vs block 1: differences from block 5 onwards).

Phasic DA responses in the ventromedial striatum quickly adapted to reversed response-reward contingencies (Fig. 2B,C). Fig. 2B shows examples of DA release in single trials with correct responses at different stages of reversal learning. Before reversal, when discrimination is well learned, DA increases to cue onset only (Fig 2B, left panel). After reversal of response-reward contingencies, cue-evoked DA signal is decreased and reward delivery now induces an increase in DA (Fig 2B, middle panel). Cue-evoked DA response is reinstated when animals have made several correct responses (Fig 2B, right panel).

Cue-evoked DA release was higher before than after reversal of response-reward contingencies (main effect reversal, $F(1,33)=28.714$, $p<0.001$; Fig. 2D, left), but did not differ between rewarded or non-rewarded trials (reversal*reward, $F(1,33)=3.789$, $p=0.06$; main effect reward, $F(1,33)=0.475$, $p=0.495$). Before reversal, reward delivery did not evoke DA release. After reversal, DA differed in rewarded and non-rewarded trials (reversal*reward interaction $F(1,33)=25.574$, $p<0.001$; reversal $F(1,33)=6.557$, $p=0.015$, reward $F(1,33)=3.415$, $p=0.074$; Fig. 2D, right). Post-hoc analysis showed that DA release at the time of lever-press increased in rewarded trials (paired t-test: $t(18)=-4.771$, $p<0.001$) but not in non-rewarded trials ($t(15)=2.334$, $p=0.034$).

For a more detailed analysis of the DA response after lever-press, we plotted changes in DA during four consecutive one-sec time bins following lever-press (Fig. 2E). We observed a decrease in DA below baseline on non-rewarded trials that was masked when analyzing the average over the four second period (time*reward interaction ($F(14,152)=9.408$, $p<0.001$); main effect time ($F(2.118,80.476)=7.201$, $p=0.001$ and reward $F(1,38)=23.435$, $p<0.001$; post-hoc analysis revealed significant effect of time for both rewarded ($F(1.091, 34.487)=7.455$, $p=0.002$) and non-rewarded trials ($F(2.212,44.243)=9.430$, $p<0.001$). The decrease in DA does not immediately follow the lever-press response, but occurs in the third and fourth second after, suggesting that the decrease in DA at a time when expected rewards are omitted occurs later than the peak increase in DA following receipt of an unexpected reward.

3.4 Trial-by-trial analysis of DA changes during reversal learning reveals rapid updating of the cue-evoked DA signal

Averaging DA traces over a complete session may obscure rapid changes in DA release patterns occurring on a trial-by-trial basis during acquisition of reversal learning (Fig. 3 lower panel; upper panel shows similar trials prior to reversal). To study the effects of positive feedback on DA release patterns, we separately analyzed the first ten trials on which the animals made a correct response after reversal (n.b. first ten *correct* trials, not always consecutive trials) and the trials that immediately followed (correct+1 trials). Our data demonstrate rapid reinstatement of cue-evoked DA release on trials that follow positive feedback. For analysis, average cue-evoked DA release on trials 1-3 was compared to average cue-evoked DA release on trials 8-10. On trials 1-3, cue-evoked DA release was higher on trials that followed a correct response (correct+1) than on trials on which the correct response was made (correct), whereas on trials 8-10 cue-evoked DA release was similar for correct trials and trials following a correct response (feedback*trial interaction ($F(1,9)=14.757$, $p=0.004$, main effect trial $F(1,9)=0.093$, $p=0.77$, main effect feedback $F(1,9)=7.795$, $p=0.021$), suggesting the effect of positive feedback on cue-evoked DA release may be most pronounced in the initial correct responses after reversal.

3.5 Learners and Non-learners differ in the extent to which reward-predicting DA signal is updated

Based on performance following reversal of response-reward contingencies, animals were divided into groups of ‘learners’(>10 total correct responses following reversal, n=11) and ‘non-learners’ (<10 correct responses following reversal, n=10; see Fig. 4A). The cut-off

criteria of 10 correct responses was based on the cumulative response curves of the animals. Our intention was to make a distinction between animals that do not learn to reverse responding at all during the reversal session and animals that are able to adapt responding to the newly rewarded side. Learners and non-learners did not differ in the amount of shaping sessions needed for lever-press training or the number of correct responses during discrimination learning. Moreover, learners and non-learners needed a similar number of trials to reach the 90% correct criterion in the discrimination phase (learners: 264.8±11.7, non-learners: 269.5±12.5 trials to criterion), suggesting the distinction was not based on a general learning defect in non-learners, but instead was specific to the reversal phase. Before reversal, response latencies on rewarded and non-rewarded trials and number of omissions were similar for learners and non-learners (Fig. 4C) suggesting motivation to respond did not differ between groups. In addition, cue-evoked DA release (Fig. 4B) was similar for learners and non-learners before reversal. After reversal, cue–evoked DA further decreased in non-learners, reflecting extinction of the conditioned response, whereas it reinstated in learners (Fig. 4B). After reversal, learners made at least 6 correct responses in a block of 10 trials. On average learners made 33.9±5.58 rewarded responses (range 16-69), whereas non-learners made 1.9±0.57 rewarded responses (range 0-6) after reversal. The number of non-rewarded responses made after reversal is similar for learners (59.1±5.5) and non-learners (55.1±7.5), but non-learners make more omissions (38.1±8.4; learners 4±2.7).

To analyze differences between learners and non-learners following reversal, we looked at the DA response for the first two correct responses (in case non-learners only made one correct response (n=3) then that value was used; excluding these animals from analysis did not affect the results) and the first two trials on which positive feedback could be used (correct+1 trials; Fig. 4E). After reversal, learners and non-learners showed similar DA release to reward delivery ($t(16)=-0.993$, $p=0.336$; Fig. 4D, Fig. 4E, left panel), but differed in cue-evoked DA release following the first correct responses (Fig. 4E, center panel). In correct+1 tria ls (Fig. 4E right panel), cue-evoked DA increased in learners, but not in non-learners. For cue-evoked DA release, we compared difference scores (peak DA value correct +1 – peak DA value correct) for the first two correct responses after reversal. The difference score was significantly higher in learners compared to non-learners ($t(11.575)=-3.851$, $p=0.002$), Fig. 4D right panel), suggesting that cue-evoked DA is updated in learners exclusively. Latency (amount of trials between reversal presentation and first correct trial) until making the first correct response after reversal did not differ between groups. Importantly, in learners, cue-evoked DA increased following trials with a correct response irrespective of latency until correct response, whereas in non-learners, cue-evoked DA was not increased after positive feedback irrespective of whether it took them longer to make the first correct response. Across all animals, we found a significant positive correlation between the percentage of correct responses after reversal and the cue-evoked peak DA response (normalized to last 10 trials before reversal to control for individual differences; $r=0.626$, $p=0.002$, Fig. 4F).

Regarding negative feedback, we compared DA release on the first two incorrect responses after reversal (error) and the trials on which this negative feedback could be used (error+1). In the first two incorrect trials, DA release after lever-press (i.e. around the time that reward was expected; Fig. 4G, left panel) did not differ between learners and non-learners

Europe PMC Funders Author Manuscripts

(t(19)=-1.475, p=0.157). Difference scores were calculated for cue-evoked DA release: Peak DA response on error trials was subtracted from peak DA response on error+1 trials and averaged across animals. No effect of negative feedback on cue-evoked DA release on consecutive trials was found (t(19)=-0.484, p=0.634; Fig. 4G, right panel). This indicates that the cue-evoked DA signal is rapidly updated following the receipt of positive feedback, but that negative feedback is not immediately reflected in the cue-evoked DA signal during reversal learning.

3.6 DA changes surrounding the time point at which learners acquire the reversal

Changes in the slope of a cumulative response record correspond to changes in performance level of the behavioral task performed (Gallistel et al. 2004). For learners, a change point was defined (trial number where a straight line drawn from origin until end of cumulative response record deviates maximally from the cumulative response curve, see Fig. 5A). The average trial for the change point was 54.9 ± 4.2 trials after presentation of the reversal. When comparing DA release on all correct responses made before and after the change point, cue-evoked DA release did not differ (paired t-test t(9)=0.420, p=0.684, Fig. 5B, left panel). However, reward delivery evoked higher DA release before the change point than after (paired t-test t(9)=3.620, p=0.006, Fig. 5B, right panel), suggesting that the switch from reward- to cue-induced phasic DA release coincided with the behavioral change point. Fig. 5C shows a quantification of the cue-evoked DA signal on trials that followed a correct response (correct+1) and trials on which the correct response was made (correct) before and after the change point. Updating of the cue-evoked DA signal after positive feedback differed before and after the change point (repeated measures ANOVA feedback*change point interaction (F(1,9)=5.278, p=0.047; main effect feedback (F1,9)=8.790, p=0.016, main effect change point F(1,9)=0.100, p=0.759). Post-hoc analysis showed that before the change point, cue-evoked DA release was higher on trials that followed a correct response (correct +1) compared to trials on which the correct response was made (correct) (t(9)=-3.278, p=0.010), suggesting that higher DA responses to reward presentation and stronger effects of positive feedback on cue-induced DA release are associated with the initial learning phase, before the behavioral change point.

4 Discussion

Successful adaptation of behavior following reversal requires the ability to use a change in reinforcing feedback. To investigate whether phasic DA release in the ventromedial striatum contributes to such an adaptation, we recorded DA release in rats during a spatial discrimination and reversal task in which a non-discriminative cue signaled trial onset. During successful responding in the discrimination phase (prior to reversal), DA was evoked by the cue, but not the reward. However, during adaptation of choice behavior following reversal of response-reward contingencies, reward delivery evoked DA release, paralleled by temporal decrease in cue-induced DA. Trial-by-trial analysis revealed rapid reinstatement of DA release to cue presentation on trials following correct responses, but no changes in DA following incorrect responses. Reinstatement of the cue-evoked DA signal was observed only in animals that learned the reversal, time-locked to their behavioral “change point”.

Europe PMC Funders Author Manuscripts

Together, this suggests that the modification of established behavior is facilitated by updating cue-evoked DA release as a consequence of positive feedback.

4.1 Phasic DA rapidly adapts to reversal of contingencies

Encountering unexpected rewards evokes a brief increase in striatal DA. After repeated pairing with a cue, the DA signal shifts from the time of reward delivery to the time of cue presentation (Day et al. 2007; Schultz et al. 1997; Pan et al. 2005), consistent with the idea that DA signaling codes a quantitative 'reward prediction error' (RPE) that serves as a teaching signal guiding behavior (Montague et al. 1996; Schultz et al. 1997; Waelti et al. 2001; Steinberg et al. 2013). Elevated DA in response to cue stimulus presentation may represent motivational properties of the stimulus and promote the initiation of reward-seeking actions (Flagel et al. 2011; Berridge et al. 2009; Wise 2004).

According to RPE theory, a reward that is fully anticipated no longer induces DA release. Consistently, we observed phasic DA release following cue onset, but not following lever-press or reward delivery during discrimination learning. During lever-press training, prior to discrimination learning, our rats learned that specific operant actions lead to reward delivery. Therefore, the shift of DA release from time of reward delivery to time of cue presentation presumably already occurred during acquisition of instrumental behavior, as shown by others (Wassum et al. 2012; Roitman et al. 2004). During discrimination learning, cue-evoked DA release did not differ on trials that followed positive feedback (correct+1 trials) and trials that followed negative feedback (error+1 trials), suggesting that when reward receipt does not differ consistently from what is expected (i.e, after lever press training), the reward-predicting DA signal is not updated on subsequent trials. Also, cue-evoked DA release was similar on trials where subjects made a rewarded response and trials where subjects made a non-rewarded response. Thus, in our task cue-evoked DA was not predictive of the subsequent choice of subject, but reflected the best available or preferred option (Roesch et al. 2007; Sugam et al. 2012). Moreover, this signal might induce incentive motivation and promote behavioral actions irrespective of trial outcome (Flagel et al. 2011; Berridge et al. 2009; Wise 2004).

Studies using long-term manipulations of the DA system (Darvas and Palmiter 2011; Clarke et al. 2011; O'Neill and Brown 2007) suggest that striatal DA contributes to the regulation of adaptive behavior. Modeling studies propose that 1) reduced DA levels after omission of an expected reward and 2) increased DA following unexpected reward or reward-predicting stimuli, may facilitate altered response execution via different basal ganglia output pathways (Hong and Hikosaka 2011; Frank and Claus 2006). Similarly, reorganization of established behavioral patterns requires suppression of DA D2-receptor mediated transmission in the nucleus accumbens, whereas acquisition and relearning of behavioral responses after a reversal or rule shift requires stimulation of accumbal D1-receptors (Yawata et al. 2012). Together, these findings suggest the importance of bidirectional phasic fluctuations in striatal DA levels when adapting behavior to changes in the environment. Although mimicking positive feedback by optogenetic stimulation of DA neurons supports reversal learning (Adamantidis et al. 2011), it is unknown whether the receipt of positive and negative

feedback during behavioral adaptation are reflected by bidirectional changes in striatal DA release on a trial-by-trial basis.

Following successful spatial discrimination, we tested the ability to modify an established response pattern after a reversal of reinforcement contingencies. Reward delivery following the initial lever presses on the newly rewarded side now rapidly increased striatal DA, as predicted by RPE theory. If DA functions as a teaching signal (Hart et al. 2014; Schultz et al. 1997) during reversal learning, the receipt of positive feedback should update the reward prediction signal in trials following correct responses. Indeed, on trials immediately following the first correct responses after reversal, cue-evoked DA increased. Thus, the receipt of positive feedback was rapidly reflected in the cue-evoked DA signal on subsequent trials, in accordance with RPE theory.

In non-rewarded trials, DA decreased below baseline after the lever press, suggesting that DA could act as a bidirectional teaching signal during reversal learning. However, this decrease was only detected across the entire reversal session (in a sec-by-sec analysis), but not in the initial incorrect responses after reversal, suggesting that this effect develops more slowly or that the decrease in DA after reward omission is relatively small, requiring a larger number of trials to be detected. This result is consistent with previously published data on extinction learning (Stuber et al. 2005; Sunsay and Rebec 2014; Owesson-White et al. 2008), but differs from the results presented by Hart et al (Hart et al. 2014). However, the latter study did not investigate the first reversal of reward contingencies, but tested extensively trained animals under frequently changing contingencies. Moreover, in our study, unexpected reward omission did not influence cue-evoked DA release on trials that immediately followed a non-rewarded response, suggesting that the receipt of negative feedback did not induce rapid updating of DA responses in our rats. However, we cannot exclude that non-reward was not sufficiently aversive to show a decrease in DA release or prolonged experience with non-reward may be needed to decrease the cue-evoked DA signal.

We show that with the completion of operant reversal learning, phasic DA in the ventromedial striatum shifts from time of reward delivery to time of cue presentation, similar to the shift of DA during initial learning of Pavlovian associations (Schultz et al. 1997; Stuber et al. 2008; Pan et al. 2005; Day et al. 2007). The effect of positive feedback on cue-evoked DA release was restricted to the reversal phase (and was not observed during the discrimination phase), corroborating previous data showing that striatal DA may be less important for learning to discriminate between two rewarded responses than for learning the reversal of such associations (Clarke et al. 2011; O'Neill and Brown 2007; Groman et al. 2011). Together, our results indicate that during the modification of established behavior, the receipt of positive feedback induces immediate updating of cue-evoked DA release, whereas the receipt of negative feedback does not. This suggests that phasic DA release during reversal learning shows an asymmetric RPE-signal (Bayer and Glimcher 2005).

4.2 Individual differences in DA signaling predict performance of reversal learning

Successful adaptation of behavior following a change in response-reward contingencies requires several processes: i) extinguish response that is no longer rewarded, ii) switch

responding to the alternative (side), iii) consolidate alternative (side) responding. These processes are thought to entail learning from both positive and negative feedback. Individual differences in sensitivity to positive feedback during reversal learning in animals have been related to D2-receptor availability (Groman et al. 2011). Similarly, in humans, learning from trial-by-trial feedback has been associated with striatal DA function (Cools et al. 2009; Frank et al. 2004; Wilkinson et al. 2014). As learning curves during reversal learning varied greatly between individuals in our study, we hypothesized that updating of cue-evoked DA release after positive feedback relates to the rate of reversal learning. Indeed, we found that animals that were slow to reverse their behavior, did not update the cue-evoked DA signal in trials following correct responses. This result has several interesting aspects. First, these non-learners were indistinguishable from learners based on behavioral and DA parameters during discrimination learning, prior to reversal.

Second, although negative feedback can robustly drive adaptation of behavior (Porter-Stransky et al. 2013), we found no difference in cue-evoked DA release to negative feedback in learners and non-learners in the initial trials following reversal, suggesting that performance differences in our reversal learning paradigm are not driven by the DA response to negative feedback.

Finally, most non-learners (8/10) sampled the reversed response-reward contingency (behavioral switch) and experienced subsequent reward delivery, but they did not sustain responding on the newly rewarded side. Instead, they returned to press the non-rewarded lever and eventually ceased lever-pressing in this session. Although we have no indication that motivation at the onset of the session was different between learners and non-learners, the increased number of omissions after reversal may indicate that non-learners differ in their motivation to respond to the newly rewarded side once the initial switch has been made.

Reward-induced DA release following the first couple of responses on the newly rewarded side could drive learning about the newly reinforced response. Moreover, the subsequent feedback-induced increase in cue-evoked DA may help to sustain motivation to regularly sample and consolidate responding to the newly rewarded side, as this was observed in learners, but not in non-learners. However, it is our experience that ‘non-learners’ generally are able to persistently switch behavior when given more time (i.e., in additional retention sessions) and that was confirmed a subset of non-learners that were exposed to more reversal sessions. This suggests that DA signaling supports a more rapid adaptation of established behavior (Klanker et al. 2013). This is similar to the presumed facilitatory, but not essential role for DA in the acquisition of reward-related learning (Robinson et al. 2005; Palmiter 2008; Darvas and Palmiter 2010; Zweifel et al. 2009).

To conclude, we showed that DA dynamics in the ventromedial striatum during reversal learning predict individual differences in adaptive behavior: Increased striatal DA following positive feedback may support the stabilization of adaptive behavior. This interpretation is further substantiated by our finding of DA changes in temporal proximity of a change point in behavior in animals that learned the reversal. Additionally, our finding that individual differences in reversal learning and associated DA release are not related to a previous

learning phase, supports the notion that these two processes are regulated by distinct mechanisms.

4.3 Conclusion

Impaired behavioral adaptation to environmental changes can result in behavioral rigidity and maladaptive behavior as observed in various neurological and psychiatric disorders, such as Parkinson's disease, drug addiction and OCD (Chamberlain et al. 2006; Cools et al. 2001; Ceaser et al. 2008; Yerys et al. 2009; Verdejo-Garcia et al. 2006). Our study suggests that individual differences in reversal learning could be related to differences in DA dynamics following positive feedback. Thus, compromised DA transmission during feedback learning could contribute to the inability to correct maladaptive behavior and the development of cognitive dysfunctions observed in psychiatric disorders such as OCD and drug addiction.

Acknowledgements

We wish to thank Ralph Hamelink and the NIN mechatronics department for technical assistance. We greatly appreciate the support offered by Drs Mark Wightman, Michael Heien, Garret Stuber, Paul Phillips and Scott Ng-Evans with the introduction of fast-scan cyclic voltammetry in our institute. This introduction was supported by two grants from the graduate school Neurowetenschappen Amsterdam.

References

Adamantidis AR, Tsai HC, Boutrel B, Zhang F, Stuber GD, Budygin EA, et al. Optogenetic interrogation of dopaminergic modulation of the multiple phases of reward-seeking behavior. J Neurosci. 2011; 31:10829–10835. [PubMed: 21795535]

Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron. 2005; 47:129–141. [PubMed: 15996553]

Bellebaum C, Koch B, Schwarz M, Daum I. Focal basal ganglia lesions are associated with impairments in reward-based reversal learning. Brain. 2008; 131:829–841. [PubMed: 18263624]

Berridge KC, Robinson TE, Aldridge JW. Dissecting components of reward: 'liking', 'wanting', and learning. Curr Opin Pharmacol. 2009; 9:65–73. [PubMed: 19162544]

Castane A, Theobald DE, Robbins TW. Selective lesions of the dorsomedial striatum impair serial spatial reversal learning in rats. Behav Brain Res. 2010; 210:74–83. [PubMed: 20153781]

Ceaser AE, Goldberg TE, Egan MF, McMahon RP, Weinberger DR, Gold JM. Set-shifting ability and schizophrenia: a marker of clinical illness or an intermediate phenotype? Biol Psychiatry. 2008; 64:782–788. [PubMed: 18597738]

Chamberlain SR, Fineberg NA, Blackwell AD, Robbins TW, Sahakian BJ. Motor inhibition and cognitive flexibility in obsessive-compulsive disorder and trichotillomania. Am J Psychiatry. 2006; 163:1282–1284. [PubMed: 16816237]

Clarke HF, Hill GJ, Robbins TW, Roberts AC. Dopamine, but not serotonin, regulates reversal learning in the marmoset caudate nucleus. J Neurosci. 2011; 31:4290–4297. [PubMed: 21411670]

Clarke HF, Robbins TW, Roberts AC. Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. J Neurosci. 2008; 28:10972–10982. [PubMed: 18945905]

Cools R, Barker RA, Sahakian BJ, Robbins TW. Mechanisms of cognitive set flexibility in Parkinson's disease. Brain. 2001; 124:2503–2512. [PubMed: 11701603]

Cools R, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J Neurosci. 2009; 29:1538–1543. [PubMed: 19193900]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Darvas M, Palmiter RD. Restricting dopaminergic signaling to either dorsolateral or medial striatum facilitates cognition. J Neurosci. 2010; 30:1158–1165. [PubMed: 20089924]

Darvas M, Palmiter RD. Contributions of striatal dopamine signaling to the modulation of cognitive flexibility. Biol Psychiatry. 2011; 69:704–707. [PubMed: 21074144]

Day JJ, Roitman MF, Wightman RM, Carelli RM. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. Nat Neurosci. 2007; 10:1020–1028. [PubMed: 17603481]

Dias R, Robbins TW, Roberts AC. Dissociation in prefrontal cortex of affective and attentional shifts. Nature. 1996; 380:69–72. [PubMed: 8598908]

Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, et al. A selective role for dopamine in stimulus-reward learning. Nature. 2011; 469:53–57. [PubMed: 21150898]

Frank MJ, Claus ED. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychol Rev. 2006; 113:300–326. [PubMed: 16637763]

Frank MJ, Seeberger LC, O'reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science. 2004; 306:1940–1943. [PubMed: 15528409]

Gallistel CR, Fairhurst S, Balsam P. The learning curve: implications of a quantitative analysis. Proc Natl Acad Sci U S A. 2004; 101:13124–13131. [PubMed: 15331782]

Groman SM, Lee B, London ED, Mandelkern MA, James AS, Feiler K, et al. Dorsal striatal D2-like receptor availability covaries with sensitivity to positive reinforcement during discrimination learning. J Neurosci. 2011; 31:7291–7299. [PubMed: 21593313]

Haluk DM, Floresco SB. Ventral striatal dopamine modulation of different forms of behavioral flexibility. Neuropsychopharmacology. 2009; 34:2041–2052. [PubMed: 19262467]

Hart AS, Rutledge RB, Glimcher PW, Phillips PE. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. J Neurosci. 2014; 34:698–704. [PubMed: 24431428]

Heien ML, Johnson MA, Wightman RM. Resolving neurotransmitters detected by fast-scan cyclic voltammetry. Anal Chem. 2004; 76:5697–5704. [PubMed: 15456288]

Hong S, Hikosaka O. Dopamine-mediated learning and switching in cortico-striatal circuit explain behavioral changes in reinforcement learning. Front Behav Neurosci. 2011; 5:15. [PubMed: 21472026]

Keithley RB, Heien ML, Wightman RM. Multivariate concentration determination using principal component regression with residual analysis. Trends Analyt Chem. 2009; 28:1127–1136.

Kim KM, Baratta MV, Yang A, Lee D, Boyden ES, Fiorillo CD. Optogenetic mimicry of the transient activation of dopamine neurons by natural reward is sufficient for operant reinforcement. PLoS One. 2012; 7:e33612. [PubMed: 22506004]

Klanker M, Feenstra M, Denys D. Dopaminergic control of cognitive flexibility in humans and animals. Front Neurosci. 2013; 7:201. [PubMed: 24204329]

Laughlin RE, Grant TL, Williams RW, Jentsch JD. Genetic dissection of behavioral flexibility: reversal learning in mice. Biol Psychiatry. 2011; 69:1109–1116. [PubMed: 21392734]

McAlonan K, Brown VJ. Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. Behav Brain Res. 2003; 146:97–103. [PubMed: 14643463]

Millan MJ, Agid Y, Brune M, Bullmore ET, Carter CS, Clayton NS, et al. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. Nat Rev Drug Discov. 2012; 11:141–168. [PubMed: 22293568]

Millar J, Stamford JA, Kruk ZL, Wightman RM. Electrochemical, pharmacological and electrophysiological evidence of rapid dopamine release and removal in the rat caudate nucleus following electrical stimulation of the median forebrain bundle. Eur J Pharmacol. 1985; 109:341–348. [PubMed: 3872803]

Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci. 1996; 16:1936–1947. [PubMed: 8774460]

O'Neill M, Brown VJ. The effect of striatal dopamine depletion and the adenosine A2A antagonist KW-6002 on reversal learning in rats. Neurobiol Learn Mem. 2007; 88:75–81. [PubMed: 17467309]

Owesson-White CA, Cheer JF, Beyene M, Carelli RM, Wightman RM. Dynamic changes in accumbens dopamine correlate with learning during intracranial self-stimulation. Proc Natl Acad Sci U S A. 2008; 105:11957–11962. [PubMed: 18689678]

Palmiter RD. Dopamine signaling in the dorsal striatum is essential for motivated behaviors: lessons from dopamine-deficient mice. Ann N Y Acad Sci. 2008; 1129:35–46. [PubMed: 18591467]

Pan WX, Schmidt R, Wickens JR, Hyland BI. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. J Neurosci. 2005; 25:6235–6242. [PubMed: 15987953]

Paxinos, G., Watson, C. The rat brain in stereotaxic coordinates. 6th edition. Elsevier Academic Press; 2007.

Phillips PE, Robinson DL, Stuber GD, Carelli RM, Wightman RM. Real-time measurements of phasic changes in extracellular dopamine concentration in freely moving rats by fast-scan cyclic voltammetry. Methods Mol Med. 2003; 79:443–464. [PubMed: 12506716]

Porter-Stransky KA, Seiler JL, Day JJ, Aragona BJ. Development of behavioral preferences for the optimal choice following unexpected reward omission is mediated by a reduction of D2-like receptor tone in the nucleus accumbens. Eur J Neurosci. 2013; 38:2572–2588. [PubMed: 23692625]

Robinson S, Sandstrom SM, Denenberg VH, Palmiter RD. Distinguishing whether dopamine regulates liking, wanting, and/or learning about rewards. Behav Neurosci. 2005; 119:5–15. [PubMed: 15727507]

Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci. 2007; 10:1615–1624. [PubMed: 18026098]

Roitman MF, Stuber GD, Phillips PE, Wightman RM, Carelli RM. Dopamine operates as a subsecond modulator of food seeking. J Neurosci. 2004; 24:1265–1271. [PubMed: 14960596]

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275:1593–1599. [PubMed: 9054347]

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci. 2013; 16:966–973. [PubMed: 23708143]

Stuber GD, Klanker M, de RB, Bowers MS, Joosten RN, Feenstra MG, et al. Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. Science. 2008; 321:1690–1692. [PubMed: 18802002]

Stuber GD, Wightman RM, Carelli RM. Extinction of cocaine self-administration reveals functionally and temporally distinct dopaminergic signals in the nucleus accumbens. Neuron. 2005; 46:661–669. [PubMed: 15944133]

Sugam JA, Day JJ, Wightman RM, Carelli RM. Phasic nucleus accumbens dopamine encodes risk-based decision-making behavior. Biol Psychiatry. 2012; 71:199–205. [PubMed: 22055017]

Sunsay C, Rebec GV. Extinction and reinstatement of phasic dopamine signals in the nucleus accumbens core during Pavlovian conditioning. Behav Neurosci. 2014; 128:579–587. [PubMed: 25111335]

Verdejo-Garcia A, Bechara A, Recknor EC, Perez-Garcia M. Executive dysfunction in substance dependent individuals during drug use and abstinence: an examination of the behavioral, cognitive and emotional correlates of addiction. J Int Neuropsychol Soc. 2006; 12:405–415. [PubMed: 16903133]

Voorn P, Vanderschuren LJ, Groenewegen HJ, Robbins TW, Pennartz CM. Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci. 2004; 27:468–474. [PubMed: 15271494]

Waelti P, Dickinson A, Schultz W. Dopamine responses comply with basic assumptions of formal learning theory. Nature. 2001; 412:43–48. [PubMed: 11452299]

Wassum KM, Ostlund SB, Maidment NT. Phasic mesolimbic dopamine signaling precedes and predicts performance of a self-initiated action sequence task. Biol Psychiatry. 2012; 71:846–854. [PubMed: 22305286]

Wightman RM, Robinson DL. Transient changes in mesolimbic dopamine and their association with 'reward'. J Neurochem. 2002; 82:721–735. [PubMed: 12358778]

Wilkinson L, Tai YF, Lin CS, Lagnado DA, Brooks DJ, Piccini P, et al. Probabilistic classification learning with corrective feedback is associated with in vivo striatal dopamine release in the ventral striatum, while learning without feedback is not. Hum Brain Mapp. 2014

Wise RA. Dopamine, learning and motivation. Nat Rev Neurosci. 2004; 5:483–494. [PubMed: 15152198]

Witten IB, Steinberg EE, Lee SY, Davidson TJ, Zalocusky KA, Brodsky M, et al. Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. Neuron. 2011; 72:721–733. [PubMed: 22153370]

Yawata S, Yamaguchi T, Danjo T, Hikida T, Nakanishi S. Pathway-specific control of reward learning and its flexibility via selective dopamine receptors in the nucleus accumbens. Proc Natl Acad Sci U S A. 2012; 109:12764–12769. [PubMed: 22802650]

Yerys BE, Wallace GL, Harrison B, Celano MJ, Giedd JN, Kenworthy LE. Set-shifting in children with autism spectrum disorders: reversal shifting deficits on the Intradimensional/Extradimensional Shift Test correlate with repetitive behaviors. Autism. 2009; 13:523–538. [PubMed: 19759065]

Zahm DS, Brog JS. On the significance of subterritories in the "accumbens" part of the rat ventral striatum. Neuroscience. 1992; 50:751–767. [PubMed: 1448200]

Zweifel LS, Parker JG, Lobb CJ, Rainwater A, Wall VZ, Fadok JP, et al. Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. Proc Natl Acad Sci U S A. 2009; 106:7281–7288. [PubMed: 19342487]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

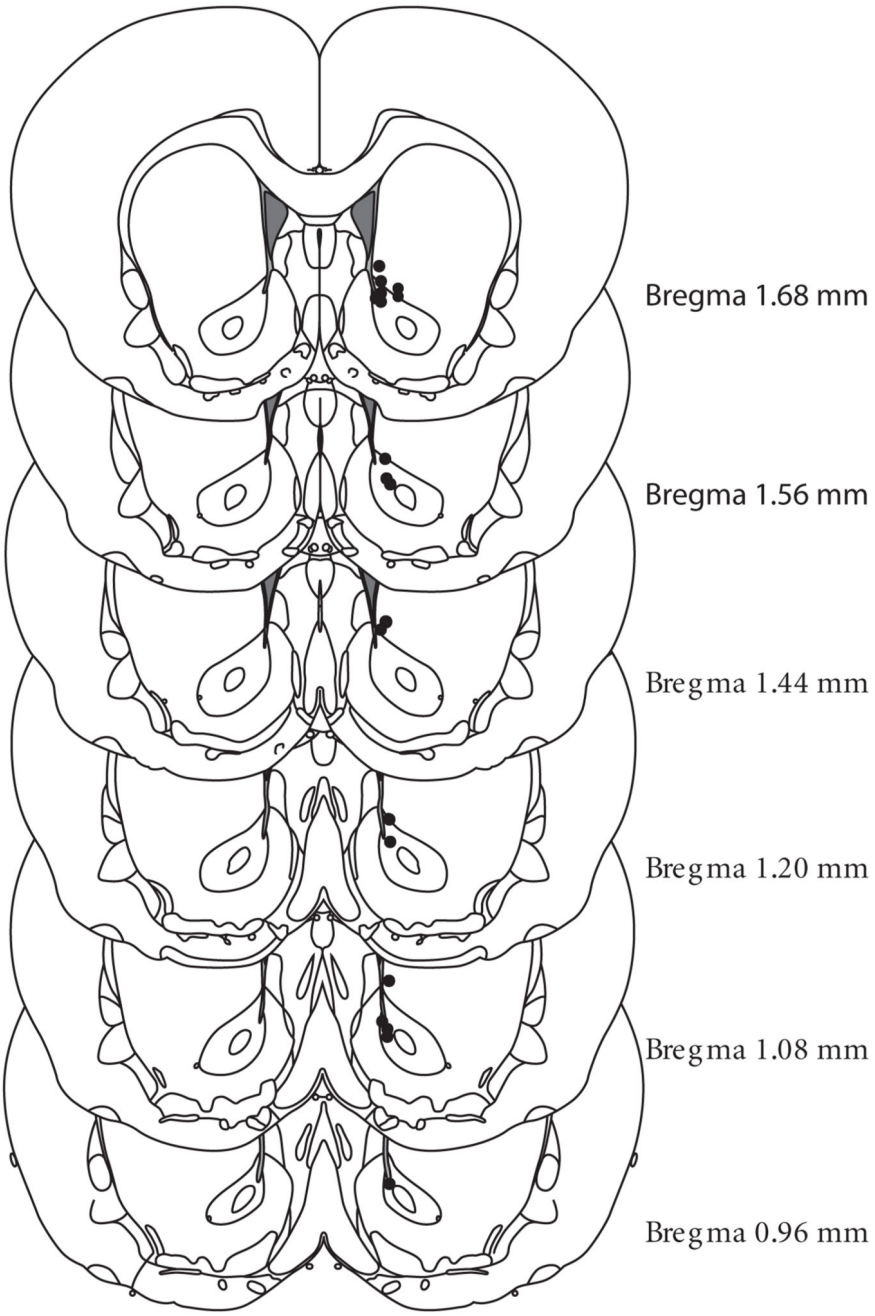

Figure 1. Histological verification of electrode placement in ventromedial striatum.
Recordings were made in nucleus accumbens core and ventromedial part of caudate nucleus right above the nucleus accumbens. Electrode placement is shown for animals, in which a post-experimental lesion was made. Each circle represents one animal.

Europe PMC Funders Author Manuscripts

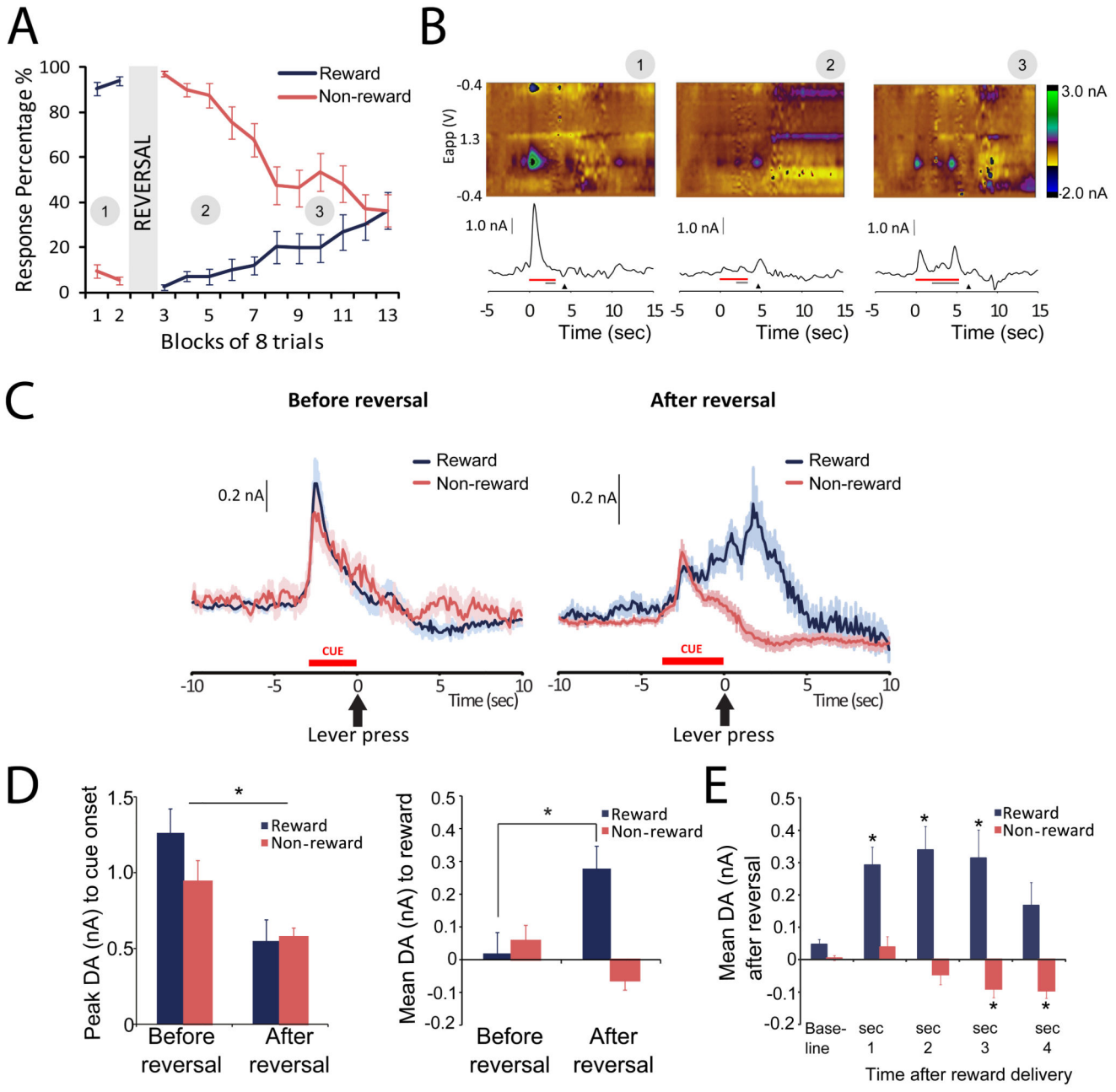

Figure 2. Phasic DA release in the ventromedial striatum during reversal learning (n=21).

A. Behavioral performance across reversal learning session. Lines show percent response during rewarded (blue) and non-rewarded (red) trials across consecutive blocks of trials before and after reversal (block 3). Numbers in grey circles correspond to examples of individual trials in panel 2B. **B.** Examples of individual trials. Red bar indicates presentation of cue lights, grey bar the presentation of levers and black triangle time of reward collection. Left – before reversal, cue presentation evokes DA release in ventromedial striatum, middle – after reversal, cue-evoked DA is diminished and reward delivery evokes DA release, right – after several correct trials, cue-evoked DA is reinstated. **C.** Fluctuations in striatal DA

averaged over trials. Left - before presentation of reversal, cue presentation evokes DA response in ventromedial striatum in both rewarded and non-rewarded trials. Right - after presentation of reversal, cue-evoked DA release is still apparent, but followed by an additional, gradual increase in DA release, in rewarded, but not in non-rewarded trials. Blue lines – mean rewarded trials, red lines – mean non-rewarded trials, shaded regions – SEM. **D.** Quantification of DA release to cue presentation and reward delivery. Left – cue-evoked DA release is lower after presentation of reversal. Right – reward delivery evokes DA release after reversal presentation. **E.** Bidirectional DA signal on rewarded and non-rewarded trials. After reversal, increased striatal DA is observed following reward delivery. In non-rewarded trials, DA decreases below baseline.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

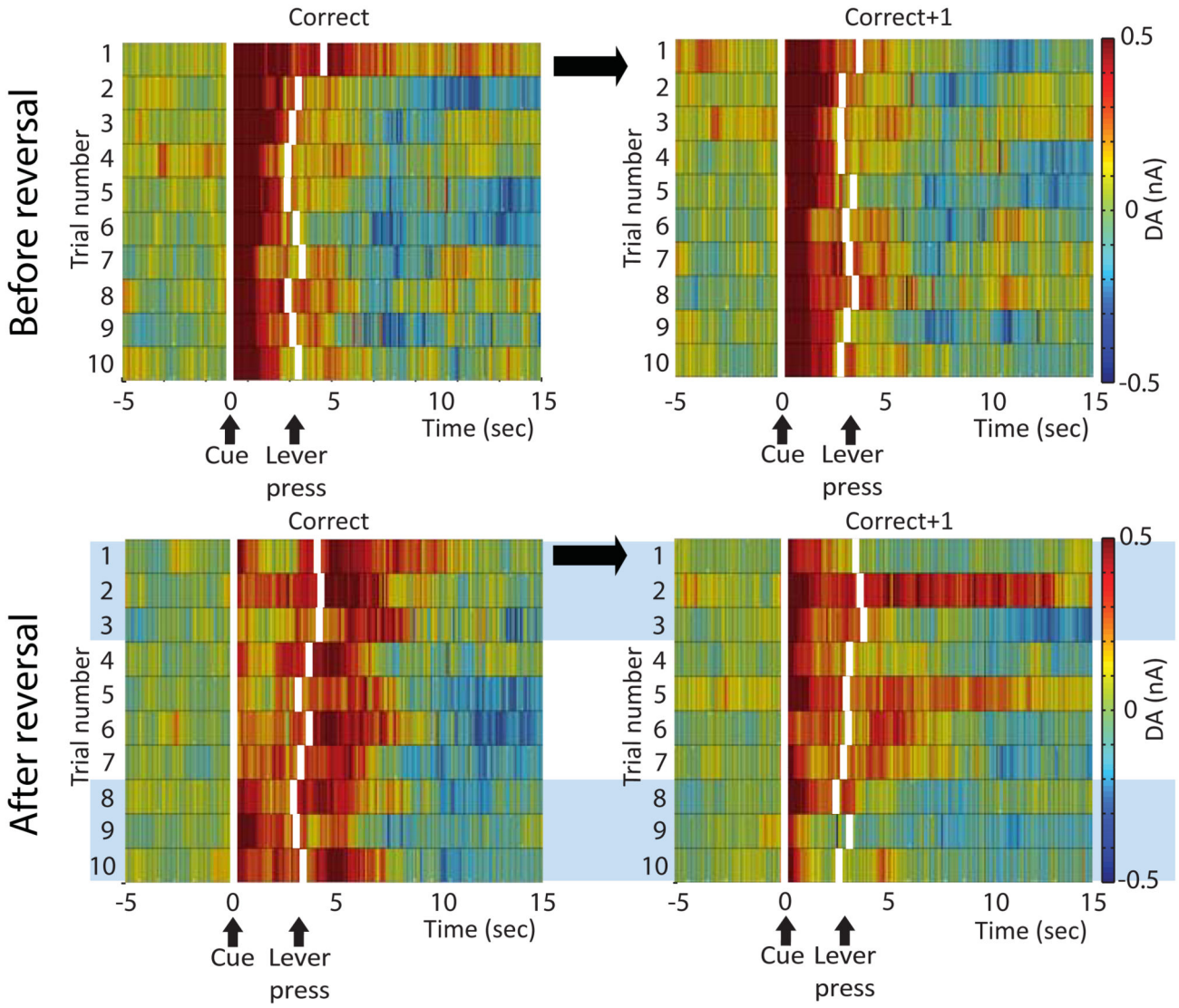

Figure 3. Rapid updating of reward-predictive DA signal after positive feedback.

Heat plots show average DA values per trial for the first 10 correct trials (correct) and the trials immediately following correct trials (correct+1) made before (top panel) and after (lower panel) reversal (shown here for animals that learned reversal, n=11). Striatal DA shows a leftward shift from time of reward delivery to time of cue presentation in first 10 trials after reversal. Receipt of unexpected reward on first correct trials after reversal induces updating of cue-evoked DA signal on trials that immediately follow the correct response. Blue boxes indicate trials used for statistical analysis. Onset of cue presentation and approximate time of lever press for each trial indicated by vertical white lines.

Europe PMC Funders Author Manuscripts

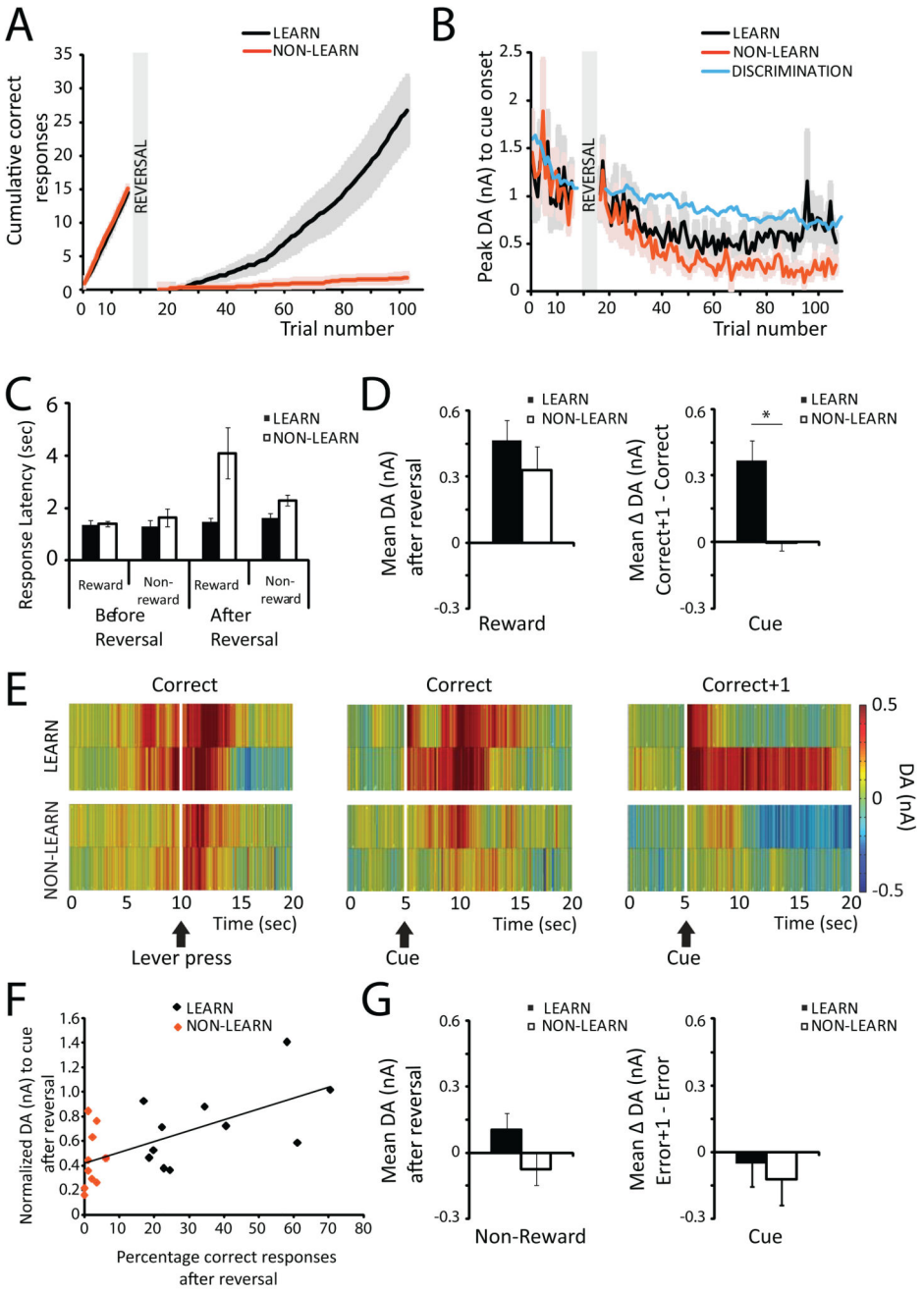

Figure 4. Individual differences in DA signaling and performance of reversal learning.

A. Cumulative response curves for learners (black, n=11) and non-learners (red, n=10). **B.** Peak values of cue-evoked DA release are similar for learners and non-learners preceding reversal presentation. After reversal, cue-evoke DA continues to decrease for non-learners, but stabilizes for learners. Learners – black, non-learners – red. Blue line shows average peak value during first session of discrimination learning for comparison.

C. Learners and non-learners show similar motivation to lever press as indicated by similar response latencies before reversal. Learners – black bars, non-learners – open bars. **D.**

Europe PMC Funders Author Manuscripts

Learners and non-learners show similar DA release to unexpected reward delivery but DA release to cues differs. Left – response to reward delivery averaged across the first two correct responses after reversal. Right – cue-evoked DA is updated after positive feedback in learners, but not in non-learners. Bar graph shows mean difference scores (cue-evoked DA on correct+1 trials – cue evoked DA on correct trials) for the first two correct responses after reversal. **E.** Changes in DA during first two correct responses after reversal. Heat plots show average DA values per trial following reward delivery (left panel) and cue presentation (center and right panels) for the first two correct trials (correct) and for the trial immediately following correct trials (correct +1). Upper panels show results for learners, lower panels show results for non-learners. Left panel – response to reward delivery, middle panel – response to cue presentation on correct trials, right – response to cue presentation on correct +1 trials. **F.** Positive correlation between percentage correct and cue-evoked DA release after reversal across all animals. Cue-evoked DA after reversal was normalized to cue-evoked DA release on last 10 trials before reversal to control for individual differences. Red dots indicate non-learners, black dots indicate learners. **G.** Changes in DA during first two incorrect responses after reversal. Left – response to lever press averaged for first two incorrect responses after reversal. Right – For cue-evoked responses difference scores between the first incorrect trials (error trials; negative feedback received) and the first trials on which this negative feedback could be used (error+1 trials) were not different between learners and non-learners.

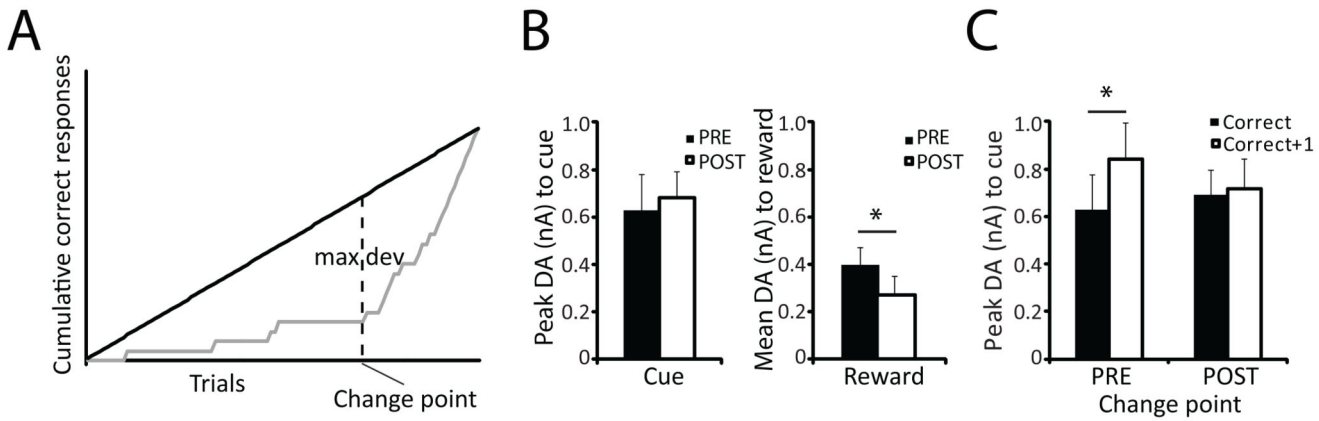

Figure 5. In animals that learned the reversal (n=11), a change point in behavioral performance is reflected in DA signal.

A. Change point in behavior was defined as the point where the cumulative correct response curve deviates maximally from a straight line drawn from the origin to the maximum of the cumulative line. **B.** Quantification of DA signal to cue presentation (left: cue) and reward delivery (right: reward) for all correct responses made before and after the change point (rewarded trials only). For rewarded trials, cue-evoked DA is similar before and after change point. DA release to reward delivery is higher before than after change point. **C.** Quantification of cue-evoked DA signal for trials on which positive feedback was received (correct) and the trials in which animals could use this feedback (correct+1) before and after the change point. Before the change point, cue-evoked DA is higher on trials that immediately follow a correct response compared to trials in which the correct response is made.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Overview procedure for behavioral training

	Lever-press training (shaping) ➡ Discrimination ➡ Reversal			
Stimuli	Randomly 1 lever + cue light		Both levers + cue lights; spatial contingency	Both levers + cue lights; contingency switch after 16-32 trials
Trials	32	64	120	120
ITI	10/20 sec	10/20 sec	15/25/35/45 sec	15/25/35/45 sec
Sessions	Until criterion	1	2*	1
Criterion for next stage	>90% response	>90% response	>90% correct responses	>90% correct responses

*additional session (max 64 tr) if criterion not reached