Different dorsal striatum circuits mediate action discrimination and action generalization

Mónica Hilario[1], **Terrell Holloway**[1], **Xin Jin**[1], and **Rui M. Costa**[1,2,*]

[1]Laboratory for Integrative Neuroscience, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland 20852-9411, USA

[2]Champalimaud Neuroscience Programme, Champalimaud Center for the Unknown, Av. Brasília, Lisbon 1400-038, Portugal

Abstract

Generalization is an important process that allows animals to extract rules from regularities of past experience and apply them to analogous situations. In particular, the generalization of previously learned actions to novel instruments allows animals to use past experience to act faster and more efficiently in an ever-changing environment. However, generalization of actions to a dissimilar instrument or situation may also be detrimental. In this study, we investigate the neural bases of action generalization and discrimination in mice trained on a lever-pressing task. Using specific schedules of reinforcement known to bias animals towards habitual or goal-directed behaviors, we confirmed that action generalization is more prominent in animals using habitual rather than goal-directed strategies. We uncovered that selective excitotoxic lesions of the dorsolateral and dorsomedial striatum have opposite effects on the generalization of a previously learned action to a novel lever. While lesions of the dorsolateral striatum impair action generalization, dorsomedial striatum lesions affect action discrimination and bias subjects towards action generalization. Importantly, these lesions do not affect the ability of animals to explore or match their lever-pressing rate to the reinforcement rate, or the ability to distinguish between different levers. The data presented here reveal that dorsolateral and dorsomedial striatal circuits have opposing roles in the generalization of previously learned actions to novel instruments, and suggest that these circuits compete for the expression of generalization in novel situations.

Keywords

habit; goal-directed; basal ganglia; motor; learning; memory

Introduction

Animals have the amazing ability to learn novel actions in order to obtain particular outcomes. Novel actions can be learned via goal-directed processes where the animal learns that a specific action leads to a particular outcome. These goal-directed actions are driven by the causal relation between the action and the outcome, and are therefore sensitive to changes in the contingency between action and outcome, and to changes in the expected value of the outcome (Dickinson *et al.*, 1983; Colwill & Rescorla, 1985; Dickinson, 1985; Balleine & Dickinson, 1994). Actions can also be learned by reinforcement of a particular stimulus-response relation; this learning process leads to the formation of habits, where the response is elicited by antecedent stimuli and is less sensitive to changes in action-outcome

*Correspondence: ruicosta@fchampalimaud.org.

relation or to changes in outcome value (Dickinson, 1985). Animals can shift between performing a similar action (e.g. driving a car in the case of humans) using goal-directed or habitual strategies. However, when faced with a novel situation or a novel instrument, animals are faced with the dilemma of having to learn a new action *de novo* or generalizing previously learned actions to the new situation/instrument (e.g. learning to drive a new vehicle that is similar to a car, versus learning to drive a new vehicle completely different to a car). Action generalization is critical because it allows animals to respond when faced with novel situations and plays a crucial role in learning behavioral policies; it would be onerous for an individual to have to learn every action *de novo* when presented with new situations/ instruments. On the other hand, individuals need to be able to discriminate when situations/ instruments are different enough and a new action/policy should be tried instead of generalizing a previous one. For example, when children learn how to bounce a ball for the first time, they may try to bounce other objects in a similar manner to see if they get a similar effect.

It has been previously shown that goal-directed actions and habits have different neural substrates (Balleine & Dickinson, 1998; Balleine *et al.*, 2003; Coutureau & Killcross, 2003; Yin *et al.*, 2004; 2005a; Yin *et al.*, 2005b; Balleine & O'Doherty, 2010). In particular, it has been shown in rats that different dorsal striatum circuits mediate goal-directed actions and habits. The dorsomedial striatum, which is connected mostly to associative areas of cortex and thalamus (Haber, 2003; Voorn *et al.*, 2004) is involved in goal-directed behavior (Yin *et al.*, 2005a; Yin *et al.*, 2005b), while the dorsolateral striatum, which is more connected to sensorimotor areas, is critical for habit formation (Yin *et al.*, 2004; 2006). However, much less is known about the neural circuits underlying the generalization of previously learned responses or the discrimination of novel situations when generalized responses may be disadvantageous.

We have previously shown using a lever pressing task that mice trained under interval schedules of reinforcement have increased predisposition towards habit formation, and that these animals generalize their responses to a novel lever never presented before (Hilario *et al.*, 2007). Conversely, mice trained on ratio schedules of reinforcement show increased tendency to learn goal-directed actions, and do not generalize their behavior towards a novel lever. The explanation put forward at the time was that if habits follow stimulus-response rules where the response is elicited by antecedent stimuli, then habitual animals would show a greater tendency to generalize towards a novel but similar lever (generalization of the response to a similar stimulus) (Hilario & Costa, 2008). On the other hand, animals displaying goal-directed actions, where the action is performed so that a particular outcome is attained, would show less tendency to generalize and would discriminate between pressing the lever in which they were trained and a novel lever. These postulates raise the possibility that the neural substrates for action generalization and habit formation are similar.

Here, we investigated the role of dorsolateral striatum (DLS) and dorsomedial striatum (DMS) in action generalization and discrimination. Firstly, we show using NMDA lesions of DMS and DLS that in mice, as in rats, DLS is important for habit formation while DMS is critical for goal-directed behavior. Secondly, we show, using a generalization test that we previously developed, that DLS is necessary for action generalization while DMS is needed for action discrimination. Finally, we show that dorsal striatum lesions do not affect the ability of animals to display matching/exploration, or the ability to distinguish between different levers. The data presented here suggests that dorsolateral and dorsomedial striatal circuits are critical for the generalization or discrimination of previously learned actions, and suggest that these circuits compete for the expression of generalization in a novel situation or in the presence of novel instruments.

Materials and Methods

Subjects

All experiments were approved by the NIAAA. Mice were purchased from Jackson laboratory at 8 weeks of age and allowed to acclimate for at least 1 week before the experiments started. One hundred and twenty nine C57BL6/J male mice between 2 and 5 months of age were submitted to surgery and used in the experiments. Only mice that learned how to press for reinforcement in the first 3 days of CRF training and mice with visible bilateral lesions in dorsomedial and dorsolateral striatum were included in the analyses. To examine the effects of dorsolateral striatum lesions forty-six animals were used. These animals were divided into 4 groups: two were trained on a random ratio schedule comprising, in one group, animals with dorsolateral striatum lesions (N=11) and, in another group, sham animals that served as controls (N=11); the other two groups were trained on a random interval schedule, which included dorsolateral lesioned animals (N=12) and their sham control group (N=12). Similarly, to investigate the effects of dorsomedial striatum lesions we used thirty animals divided into 4 groups: two groups were trained on a random ratio schedule including animals with dorsomedial striatum lesions in one group (N=7) and their sham controls (N=8); while the other two groups were trained on a random interval schedule, which included dorsomedial lesioned animals (N=8) and their corresponding sham animals (N=7).

Stereotaxic surgery and lesions

Stereotaxic neurosurgery was conducted under general isofluorane anesthesia. Excitotoxic lesions of the striatum were performed by injecting NMDA into the striatum using a syringe pump (Razel Scientific). Two cannulae were inserted per striatum (two sites per dorsolateral or dorsomedial region on each hemisphere, one more anterior and one more posterior, along a line connecting +1.18 mm anterior-posterior, +2.0 mm medial-lateral with +0.22 mm anterior-posterior, +2.6 mm medial-lateral). We injected 0.25–0.4 μl of a 10 mg ml^{-1} solution of NMDA per site (2.25 mm below the surface of the brain) at a rate of 5–10 μl h^{-1}. The cannulae were left in place for an extra 15 minutes to allow for drug diffusion. Control mice had cannulae inserted, similarly to the lesioned mice, but no drug was delivered. Mice were allowed to recover for two weeks before behavioral training started. Lesions were verified postmortem using Nissl staining of 50 μm brain slices, after perfusion and overnight postfixation with 4% paraformaldehyde (by weight) (Supplementary Figure 1).

Behavioral Procedure

The apparatus and training procedures were identical to those used in a previous study (Hilario *et al.*, 2007). Behavioral training and testing took place in operant chambers (21.6 cm L × 17.8 cm W × 12.7 cm H) housed within sound attenuating chambers (Med-Associates, St. Albans, VT). Each chamber was equipped with two retractable levers on either side of a food magazine, and a house light (3W, 24V) mounted on the opposite side of the chamber. Reinforcers were delivered into the magazine through a pellet dispenser or a pump with a syringe that delivered sucrose solution (20–30 μl of 10% solution per reinforcer). Magazine entries were recorded using an infra-red beam and licks using a contact lickometer. Before training started mice were placed on a food deprivation schedule, receiving 1.5 to 2g of food per day allowing them to maintain a body weight above 85% of their baseline weight. Throughout the training phase mice were fed daily after the training session. Water was removed for 4–6 hours before each daily session. Animals were trained with two reinforcers: pellets (Bio-Serv– formula F05684 or sucrose 20 mg) or sucrose (10%). One reinforcer was delivered in the operant chamber contingent upon lever pressing, and the other reinforcer was presented freely in their home cage and used as a control for the devaluation test. The reinforcer and lever used were counterbalanced across animals.

Training started with a 30 minute magazine training session in which one reinforcer was delivered on a random time schedule on average every 60 s (30 reinforcers). The following day lever pressing training started, in which each animal learned to press one lever to obtain a specific reinforcer. Each daily session began with the illumination of the house light and insertion the lever and ended with the retraction of the lever and the offset of the house light. Typically, lever pressing training commenced with 3 sessions of continuous reinforcement (CRF) in the first 3 days. The first CRF sessions lasted 90 minutes or until the mice received 5 reinforcers, the second CRF session lasted 90 minutes or until the mice received 15 reinforcers, and the last CRF session lasted 90 minutes or until the mice received 30 reinforcers. After CRF, animals were trained in either ratio or interval schedules, with all the sessions lasting 90 minutes or until mice received 30 reinforcers. For random ratio training, after the last session of CRF, mice were given one session of random ratio 10 (RR-10) and then switched to random ratio 20 (RR-20; on average one reinforcer every 20 lever presses). For interval training, after CRF mice were given one session of random interval 30 (RI-30) and then switched to random interval 60 (RI-60 –reinforcer delivered upon the first lever press after 60 seconds on average elapsed 10% probability of reinforcer delivery every 6 seconds).

The devaluation test lasted 2 days. On each day mice were given ad libitum exposure to one of the reinforcers for 1 hour in a separate cage. Mice were allowed to consume either the reinforcer earned by lever pressing (devalued condition), or the reinforcer they received for free in their home cage (valued condition). The amount of reinforcer consumed during the ad libitum session was recorded, and mice that did not consume a minimum of 0.5g of each reinforcer were not included in the analyses. Immediately after the ad libitum feeding session, mice were given a 5 min extinction test with the training lever extended. The number of presses on the training lever under the valued and the devalued conditions was compared.

The generalization test consisted of a 5 minute choice test in which two levers were presented - the lever on which the animals were trained and a novel lever. The number of presses on each lever was recorded. The test was given in extinction and not preceded by feeding.

For the matching experiments, mice were submitted to two daily training sessions for 4 consecutive days. Each day, animals underwent two separate sessions, and in each of the sessions they were exposed to only one lever (forced choice). Each lever was associated with a different reinforcement rate. Random ratio trained animals were trained with RR-20 on one lever and RR-30 on the other - designated as ratio higher and lower reinforcement rate levers, respectively. Random interval trained animals were trained RI-60on one lever and RI-90 on the other – designated as interval higher and lower reinforcement rate levers, respectively.. Forced-choice training sessions were terminated upon the completion of one of the two criteria: the earning of 15 reinforcers or the end of the time allocated for the training session (60 min). All animals were able to earn the 15 reinforcers before 60 minutes had elapsed. The order of the forced-choice training sessions for each lever was counterbalanced across days, and training on each of the levers had an inter-session interval of four hours.

Following forced-choice training, animals were tested on a choice test with both levers present - free choice test. This free choice test lasted 5 minutes and was performed in extinction).

Statistics

For all sessions, the number and time of lever presses, head entries, and reinforcements were recorded. Statistical Analyzes were done using SPSS. Data relative to training measurements and cumulative lever pressing were analyzed using Repeated Measures Analysis of Variance (ANOVA). *Post hoc* analyses were performed using Bonferroni. Since our experiments were designed to compare specifically two conditions per group of animals, the devaluation and generalization data were analyzed using planned comparisons employing two-tailed paired t-tests, with the null hypothesis being that there were no statistical difference between conditions, and the alternative hypothesis being that the two conditions were different. Nonetheless, we also performed two-way analysis of variance on all the devaluation data with the factorial design (reward value (2 levels: valued/devalued) × lesions and schedules (8 levels: DLS NMDA/SHAM lesioned animals RI trained/RR trained and DMS NMDA/SHAM lesioned animals RI trained/RR trained) and observed a significant interaction between variables ($F_{7,61}$ =2.47, p=0.026) and a significant effect of reward value ($F_{1,61}$=18.25, p<0.0001). For the generalization data, a two-way analysis of variance with the factorial design (reward value (2 levels: hard/easy) × lesions and schedules (8 levels: DLS NMDA/SHAM lesioned animals RI trained/RR trained and DMS NMDA/SHAM lesioned animals RI trained/RR trained) yielded a significant effect of lever ($F_{1,68}$=44.0, p<0.0001, albeit no significant interaction $F_{7,68}$ =0.86, p=0.5452). The rate of responding during testing was normalized to the rate of responding during the last day of training for each animal. Significance was set to α = 0.05 for all tests performed. Figures 1,2,3 and Supplementary Figure 2 show group means and standard error of the mean (SEM). Figure 4 shows correlations between lever presses and reinforcers across the different animals and across the different groups, in the forced choice and the free choice conditions. Correlations were performed using Pearson correlation.

RESULTS

Dorsolateral and dorsomedial striatum are differentially involved in habits and goal-directed actions in mice

We previously reported that in mice (Hilario *et al.*, 2007), as in other species (Dickinson *et al.*, 1983; Yin *et al.*, 2004; Balleine & O'Doherty, 2010), ratio schedules of reinforcement on operant tasks favor the learning and execution of goal-directed actions while interval schedules of reinforcement favor habit formation. Studies in rats have shown that goal-directed actions and habits are dependent on different dorsal striatal circuits (Yin *et al.*, 2004; 2005a; Yin *et al.*, 2005b; Yin *et al.*, 2006). We thus examined if the development of goal-directed actions and habits in mice trained in different schedules of reinforcement was also dependent on dorsomedial and dorsolateral striatum. We trained dorsolateral and dorsomedial striatum lesioned mice and respective sham controls on different schedules of reinforcement (in total eight experimental groups) in an operant task where animals had to press one lever to obtain a particular outcome. To analyze the effects of dorsolateral lesions we compared four groups of mice: two groups trained on a random interval schedule - dorsolateral lesioned animals (DLS Lesions interval) and their controls (DLS Sham interval); and other two groups trained on a random ratio schedule -dorsolateral lesioned animals (DLS Lesions ratio) and their controls (DLS Sham ratio). Similarly, the effects of dorsomedial striatum lesions were examined using four groups of mice: two groups trained on a random interval schedule - dorsomedial lesioned animals (DMS Lesions interval) and their controls (DMS Sham interval); and two other groups trained on a random ratio schedule - dorsomedial lesioned animals (DMS Lesions ratio) and their controls (DMS Sham ratio) (Supplementary Figure 1).

During training, all animals from the different DLS groups showed increased lever pressing ($F_{6, 132} = 37.9$, p<0.001), with a significant interaction between the groupsand training ($F_{27, 378} = 4.06$, p<0.0001), and a significant difference between groups ($F_{3,378} = 5.54$, p=0.003) (Fig. 1A). Bonferroni's *post hoc* analyses revealed no differences between sham and DLS lesioned animals trained on a random ratio or between sham and DLS lesioned animals trained on a random interval schedule. However, as expected, significant differences were found between the sham ratio and sham interval groups during the last 5 days of training and between lesioned animals trained on a random ratio and random interval schedule during the last 2 days of training ($p < 0.05$). Similarly, animals from the DMS groups showed increased lever pressing across time ($F_{9,234}= 50.9$, p<0.0001), and a significant interaction ($F_{27,234}= 3.08$, p<0.0001). Bonferroni's *post hoc* analyses found no difference between sham and lesioned animals, but revealed a significant difference between shams trained on a random interval and random ratio schedule during the last 4 days of training, as well as between lesioned groups trained on different schedules ($p < 0.05$). These results are congruent with previous findings showing that ratio schedules lead to more lever pressing than interval schedules (Dickinson *et al.*, 1983; Hilario *et al.*, 2007), but reveal that neither DLS nor DMS lesions affect lever pressing during training. Furthermore, magazine exploration, as measured by headentries into the magazine, was also not affected in lesioned animals (Fig. 1B). All DLS animals increased head entries throughout training ($F_{9,378}= 6.92$, p<0.0001), and there was no significant effect of training schedules or lesions ($F_{3,378}= 1.20$, p=0.3198), and no significant interaction ($F_{27,378}= 0.77$, p=0.79). Similarly, DMS groups also showed increased head entries with time ($F_{9,234}= 8.56$, p<0.0001) but no interaction effect ($F_{27,234}= 0.46$, p=0.99) or differences between groups ($F_{3,234}= 1.07$, p=0.38).

In order to investigate if selective NMDA-lesions of DLS and DMS affected the ability of animals to behave in a habitual or goal-directed manner we performed a devaluation test. Consistently with previous results (Hilario *et al.*, 2007), DLS sham animals trained on a random ratio schedule responded significantly less during the devalued condition, when the outcome they pressed for during training was devalued by sensory-specific satiety, than during the valued condition when they were pre-fed a control outcome ($T_{10} = 4.68$, p = 0.0009), indicating that they behaved in a goal-directed manner. Conversely, the random interval DLS sham group pressed similarly during the valued and devalued conditions ($T_{11} = 0.12$, p = 0.91), suggesting that their responding was habitual i.e. insensitive to changes in the expected value of the outcome. However, both groups of dorsolateral striatum lesioned mice showed sensitivity to sensory specific satiety, pressing significantly more during the valued condition, independent of the schedule they were trained on (ratio trained animals $T_{10} = 3.72$, p = 0.004; interval-trained animals $T_{11} = 3.44$, p = 0.005, Fig. 1C), indicating that DLS in mice is critical for habit formation.

In accordance with the results presented above, DMS sham ratio-trained animals showed significant devaluation ($T_7 = 3.21$, p = 0.015), whereas DMS sham animals trained under interval schedules became habitual and did not show an effect of devaluation ($T_5 = 0.99$, p = 0.36). Nonetheless, both DMS lesioned mice trained under interval schedules ($T_4 = 0.37$, p = 0.73) and under ratio schedules ($T_3 = 0.09$, p = 0.934) failed to show sensitivity to sensory specific satiety (see Fig. 1C), indicating that DMS is critical for goal-directed behavior.

Overall, these results corroborate previous findings (Yin *et al.*, 2004; 2005a; Yin *et al.*, 2005b; Yin *et al.*, 2006) by showing that in mice DMS is critical for goal-directed actions, while DLS is necessary for habit formation.

Dorsolateral striatum is critical for action generalization

We next investigated how dorsal striatum lesions would affect the ability of animals to generalize previously learned actions to novel but similar instruments. To assess this, we

used a generalization test that we previously developed (Hilario *et al.*, 2007). This generalization test consisted of a two-lever choice test, where animals had to choose between pressing the lever that was reinforced during training, or a novel identical lever that was presented in another location of the operant chamber.

In accordance with our previous findings, DLS sham animals trained on random ratio-schedules pressed significantly more on the trained than on the novel lever (Fig. 2A, $T_{10} = 4.66$, p = 0.001; Fig. 2B, cumulative pressing on each lever across the five minutes of the test, $F_{1,580}=6.45$, p=0.019), whereas the DLS sham animals that were trained on random interval schedules pressed the novel lever as much as the training lever (Fig. 2A, $T_{11} = 1.65$, p = 0.13; Fig. 2B, $F_{1,754}=0.24$, p=0.63). These data indicate that random interval trained animals generalize a previously learned action to a new lever presented during a novel situation, while random ratio trained animals discriminated between the two actions and did not generalize the previously learned action. Interestingly, dorsolateral lesioned mice trained in both schedules pressed the novel lever significantly more than the trained lever - ratio trained animals ($T_{10} = 3.14$, p = 0.01, $F_{1,638}=6.20$, p=0.021); interval-trained animals ($T_{11} = 3.46$, p = 0.005, $F_{1,580}=6.96$, p=0.016) (Fig. 2A, 2B). These results suggest that the dorsolateral striatum is critical for the generalization of a previously learned response in a novel situation.

Dorsomedial striatum is necessary for action discrimination

We also examined the effects of DMS lesions on the generalization test. DMS sham animals trained on random ratio-schedules pressed significantly more of the trained than on the novel lever ($T_7 = 2.65$, p = 0.033, $F_{1,406}=5.70$, p=0.0317), whereas the animals that were trained on random interval schedules pressed the novel lever as much as the training lever ($T_6 = 2.469$, p = 0.05, $F_{1,348}=2.02$, p=0.181, Fig. 3A, B). However, dorsomedial lesioned mice that were ratio-trained ($T_6 = 0.89$, p = 0.40, $F_{1,348}=0.69$, p=0.42) and interval-trained ($T_7 = 1.89$, p = 0.10, $F_{1,348}=0.62$, p=0.45) pressed the novel lever as much as the trained lever (Fig. 3A, 3B), indicating that DMS lesioned animals have a higher tendency for generalization, and are unable to discriminate actions on the training from actions on the novel lever. These results also suggest a competition between dorsomedial and dorsolateral striatal circuits in action generalization and discrimination.

Dorsal striatum lesions do not affect lever discrimination or matching/exploration

The observation that during the generalization test mice trained on interval schedules pressed the novel lever similarly to the training lever could be attributed to a generalization of the stimulus-response relation to a novel action. However, since in interval schedules time has to pass between the availability reinforcements, it is less costly for interval-trained animals to spend time pressing a novel lever. Indeed, interval schedules have been shown to favor exploratory behavior and matching (Herrnstein, 1961; Staddon & Cerutti, 2003), and so it plausible to consider that animals trained on an interval schedule could be checking the new lever as part of an exploratory/matching strategy. Another possibility to consider is that some of the effects of the lesions reported above could stem not from an inability of discriminating or generalizing the actions to perform on the different levers, but from effects of the lesions on the ability to distinguish the two levers.

In order to clarify if the effects observed above could result from changes in exploration/matching, or changes in the ability to distinguish the two levers, we performed additional experiments. We trained each animal on 2 levers to obtain the same outcome, with each lever associated with a different reinforcement rate. Animals previously trained on random ratio using an RR-20 schedule on one lever, were trained with RR-20 on one lever (higher reinforcement rate) and RR-30 on another lever (lower reinforcement rate). Mice were first

trained for 4 days with two subsequent sessions a day, one on each lever and counterbalanced (forced-choice), and then tested on a choice test with both levers present (free choice, test lasted 5 minutes and was performed in extinction). Mice previously trained on random interval using an RI-60 schedule on one lever, were trained with RI-60 on one lever (higher reinforcement) and RI-90 on another lever (lower reinforcement). As in the random ratio case, mice were first trained for 4 days with two subsequent sessions a day, one on each lever, and then tested on a choice test with both levers present.

The forced-choice experiment allowed us to investigate if mice can match their response to different reinforcement rates, independently of the schedule used. We found that during forced-training, when only one schedule/lever was presented per session, mice from the different groups matched their response rate to the reinforcement rate very well (Fig. 4A, R^2 >0.9, right panels for both DLS and DMS). This matching was also found when considering all animals independently, and no difference was found between sham animals and lesioned animals for both DMS and DLS (Fig. 4A, left panels for both DLS and DMS). These results indicate that, during the forced-choice training, mice matched their responses to the reinforcement rate, and there were no effects of lesions on matching ability. During the free-choice test, although across all groups there was a significant correlation between the choice rate of each lever and the reinforcement rate of that lever during forced-choice training (Fig. 4B, right panels), interval and ratio-trained animals behaved differently.. While we observed that interval-trained animals matched their response rate on each lever during test to the reinforcement rate they received during training, ratio trained animals showed an exploitative pattern of behavior, as they chose significantly less the lower reinforcement lever than predicted by matching (Fig 4B, Supplementary Figure 2 B). Still, no effects of lesions were observed during the free-choice test (Fig 4B left panels). Taken together, these data show that neither DMS nor DLS lesions affect the ability of animals to explore/match or exploit, and suggest that action generalization/discrimination and matching/exploitation emerge from dissociable neural circuits.

The data presented above also suggests that lesioned animals can distinguish between the two levers and the reinforcement rates associated with them. To more formally examine this, we used two-way repeated measures ANOVAs, with lever difficulty as the repeated measures, to look for differences between performance during the 4 days of forced-choice training of sham and lesioned animals. During training, there were no significant differences in lever pressing between random-ratio trained DLS sham and DLS lesioned groups ($F_{1,20}$ = 0.55, p=0.47). There was also no interaction between levers and lesions ($F_{1,20}$ = 2.39, p=0.14); however, both lesioned and sham animals pressed the higher reinforcement lever significantly more than the lower reinforcemnt lever ($F_{1,20}$=32.63, p<0.0001) (Supplementary Figure 2). Random-interval trained DLS sham and DLS lesioned animals also did not differ in lever pressing ($F_{1,21}$ = 0.00, p=0.98), and no significant interaction was observed between lever choice and lesion ($F_{1,21}$ = 0.12, p=0.73). Similarly, random-ratio trained DMS sham and DMS lesioned groups did not differ in their lever pressing ($F_{1,13}$ = 1.58, p=0.23), and both groups pressed the higher reinforcement lever significantly more ($F_{1,13}$ = 9.45, p=0.01), with no interaction between lever pressing and lesion ($F_{1,13}$ = 0.01, p=0.91). Consistently, random-interval trained DMS sham and DMS lesioned groups also did not show a difference in lever pressing ($F_{1,13}$ = 0.03, p=0.87), or an interaction between lever pressing and lesions ($F_{1,13}$ = 0.13, p=0.73), but both sham and lesioned groups pressed higher reinforcement lever significantly more ($F_{1,13}$ = 6.66, p=0.03). As expected, during forced-choice training, a significant higher reinforcement rate was observed in all groups for the lever requiring lower effort (DLS Sham and lesions trained on a random interval, $F_{1,42}$=90.5, p<0.0001; DLS Sham and lesions trained on a random ratio, $F_{1,40}$=68.6, p<0.0001; DMS Sham and lesions trained on a random interval, $F_{1,26}$=44.8, p<0.0001; DMS

Sham and lesions trained on a random ratio, $F_{1, 26}=6.06$, p=0.02); however no significant differences between groups or interactions were found (Supplementary Fig. 2 A).

We also performed similar analyses for the free-choice test (Supplementary Fig. 2 B). DLS interval-trained shams and lesioned animals chose more often the lever with higher reinforcement rate during training ($F_{1,21} = 11.40$, p=0.0028), with no effect of the training schedule ($F_{1,21} = 0.34$, p=0.57), and no significant interactions among factors ($F_{1,21} = 0.01$, p=0.93). Additionally, DLS random ratio shams and lesioned animals preferred the higher reinforcement lever ($F_{1,20} = 101$, p<0.0001), with no effect of training schedule ($F_{1,20} = 0.77$, p=0.39), and no significant interactions among factors ($F_{1,20} = 2.32$, p=0.1432). DMS random interval shams and lesioned animals also preferred the higher reinforcement lever ($F_{1,12} = 11.1$, p<0.006), with no effect of training schedule ($F_{1,12} = 0.06$, p=0.80), and no significant interactions among factors ($F_{1,12} = 2.06$, p=0.18). Finally, DMS random ratio shams and lesioned animals also chose to press more frequently the higher reinforcement ($F_{1,13} = 76.7$, p<0.0001), with no effect of the training schedule ($F_{1,13} = 0.81$, p=0.39), and no significant interactions among factors ($F_{1,13} = 0.00$, p=0.99) (Supplementary Fig. 1B).

These data indicate that DLS and DMS lesions do not affect the ability of animals to explore different levers, to match their response rate to the reinforcement rate, or to distinguish between the two levers. Therefore, the results obtained in the generalization test are unlikely to be explained by differences in matching in DLS lesioned interval-trained animals, or differences in the ability to exploit in DMS lesioned ratio-trained mice. Rather, they are likely to reflect differences in the ability to generalize a previously learned action to a novel situation in the case of DLS lesioned animals, or the ability to discriminate and not generalize a previously learned action to a novel situation in the case of DMS lesioned animals.

DISCUSSION

We investigated the role of dorsomedial and dorsolateral striatum in action generalization by training animals in different reinforcement schedules known to promote different predisposition for habit formation and action generalization. Our results show that, as previously observed in rats (Yin *et al.*, 2004; Yin *et al.*, 2005b), mice with specific excitotoxic lesions of the dorsolateral striatum are impaired in habit formation but have intact goal-directed behavior. Importantly, we show that unlike sham controls trained in interval schedules of reinforrocement, mice with excitotoxic lesions of DLS do not generalize their responses to a novel lever. Conversely, we confirmed that dorsomedial striatum lesions in mice rendered the animals incapable of goal-directed behavior without affecting habit formation. Furthermore, mice with DMS lesions displayed a higher tendency for action generalization and lack of discrimination between the training lever and a novel lever, suggesting that DMS and DLS compete, and that in DMS lesioned animals the intact dorsolateral striatum controls behavior. These results are congruent with previous studies (Hilario *et al.*, 2007), where we performed a similar devaluation test followed by a generalization test in mice carrying a homozygotic null mutation of the cannabinoid receptor type I (CB1), which is highly expressed in dorsolateral striatum. CB1 knockout mice showed impaired habit formation and a reduced level of generalization..

The increased pressing of a novel lever in interval schedules of reinforcement compared to ratio schedules of reinforcement could stem from an enhanced predispostion for exploration or matching in interval schedules of reinforcement (Herrnstein, 1961; Staddon & Cerutti, 2003), rather than from generalization of the previously learned response. We therefore designed an experiment where the two levers would lead to the same outcome but with different schedules of reinforcement. We found that overall mice matched their lever

pressing rate to the reinforcement rate, and that neither lesions of dorsomedial striatum nor lesions of dorsolateral striatum affected the ability of mice to match. Furthermore, even when ratio-trained animals displayed exploitative behavior during a choice test, DMS or DLS lesions did not affect this behavior. These tests also showed that lesioned mice are able to distinguish between the two levers. These data therefore attest the specificity of our findings. Furthermore, our results argue for functional and anatomical independence between exploratory/matching and generalization, which is consistent with previous studies which identified the frontopolar, medial frontal and posterior parietal cortex as areas activated during exploration (Daw *et al.*, 2006).

Our findings extend our knowledge on the role of the striatum in rule generation and generalization. The cortico-basal ganglia system is composed of independent parallel, but interacting, corticostriatal loops specialized in processing different types of information (Haber, 2003; Voorn *et al.*, 2004). The connections between the cortical and thalamic regions and the striatum generate anatomical and functional ventrodorsal and mediolateral directional gradients that reflect differences in the content of learning, memory consolidation and automatization, for example. Dorsomedial striatum has been shown to receive most of its input from the associative areas of the cortex and to support learning and performance of goal-directed behavior, while the dorsolateral striatum has been shown to receive input from the sensorimotor areas of the cortex and to support the formation of habits (Voorn *et al.*, 2004). Other studies have shown that the dorsomedial striatum is important for the early phases of skill learning, while dorsolateral striatum starts to be engaged early but becomes more engaged as skills are consolidated (Yin *et al.*, 2009). It has also been shown that the caudate (roughly homologous to the DMS) seems to be preferentially involved in the initial stages of visuomotor learning while the putamen (roughly homologous to DLS) seems to be more critical when behavior is automated (Miyachi *et al.*, 1997; Miyachi *et al.*, 2002). Additionally, it has been shown that neural activity related to the start/stop of motor sequences emerges preferentially in dorsolateral striatum compared to dorsomedial striatum (Jin & Costa, 2010), and that the dorsolateral but not the dorsomedial striatum is important for serial order learning (Yin, 2010). All of the studies mentioned above support a differential functional role of DMS and DLS in the learning of motor actions, a finding that is corroborated by our results showing differential involvement of these striatal circuits in action generalization/discrimination.

Interestingly, the dorsomedial and dorsolateral striatum have also been shown to be differentially involved in place and response learning, respectively (Yin & Knowlton, 2004), which suggests that the classical theory that basal ganglia and hippocampal memory systems interact competitively (Packard & McGaugh, 1992; Poldrack & Packard, 2003; Atallah *et al.*, 2004; Squire, 2004; Seger & Cincotta, 2005), may be more complex than previously acknowledged. Previous studies showed that with the passage of time as memories becomes less dependent on the hippocampus they become more general and schematic (Wiltgen & Silva, 2007; Wiltgen *et al.*, 2010). Given the role of both dorsomedial striatum and hippocampus in place learning, and the extensive anatomical links between the hippocampus and dorsomedial striatum (Finch *et al.*, 1995), it would be interesting to examine if the differential role in action discrimination and generalization reported here for DMS and DLS could be applied more generally to memory generalization and to categorization. Categorization requires some level of generalization, and the potential role of specific corticostriatal loops in categorization has recently received more attention (for a review, see (Seger, 2008). Studies relating language rule application and basal ganglia disorders such as Parkinson's and Huntington's disease have also started to emerge. Patients with Parkinson's disease show greater impairment of their native language, acquired at early ages, compared to languages learned later in life (Zanini *et al.*, 2010). It was proposed that this is because components of native language that are described in terms of rules (phonology, morphology,

syntax, and the morphosynthetic properties of lexicon) are more dependent on fronto-striatal circuits (procedural implicit processing), while other languages learned at later times are encoded by temporo-parietal structures (declarative memory processing) (Ullman, 2001; Zanini *et al.*, 2010). Similarly, studies on patients with Huntington's disease, which leads to neuronal death in the dorsal striatum, found that patients have impaired ability to transfer acquired structural knowledge from one language to a different language (De Diego-Balaguer *et al.*, 2008). Additionally, recent studies suggest a role for corticostriatal interactions on the impaired ability to generalize observed in patients with autism spectrum disorders (Solomon *et al.*, 2011). Nevertheless, further studies are needed to identify the molecular mechanisms supporting action generalization, and to clarify if the role of dorsolateral striatum in action generalization can be extended to rule-application, categorization, and language processes.

In conclusion, the results presented here suggest that dorsolateral and dorsomedial striatal circuits are differentially involved in the generalization of previously learned actions to a novel instrument. These data suggest that different cortico-basal ganglia circuits may compete during novel situations when organisms need to decide which action policy to adopt. These results may have implications for understanding rule learning and categorization deficits observed in basal ganglia disorders.

Supplementary Material

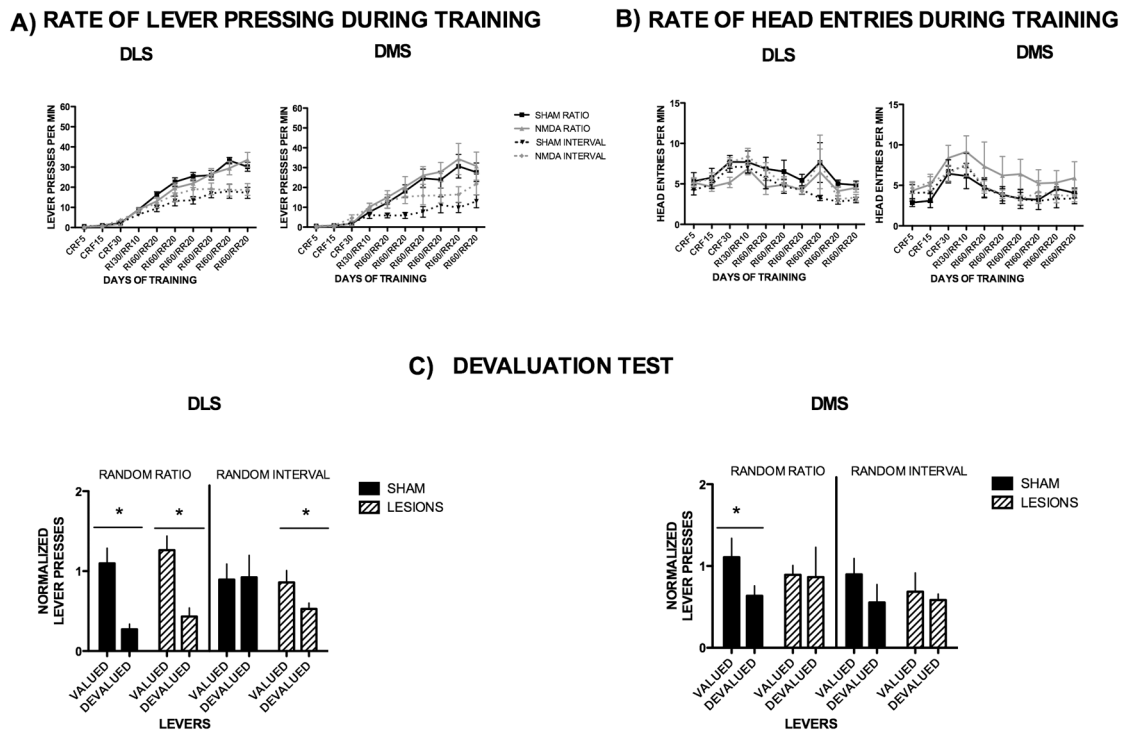Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Atallah HE, Frank MJ, O'Reilly RC. Hippocampus, cortex, and basal ganglia: insights from computational models of complementary learning systems. Neurobiol Learn Mem. 2004; 82:253–267. [PubMed: 15464408]

Balleine BW, Dickinson A. Motivational control of goal directed action. Animal Learning and Behavior. 1994; 22:1–18.

Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology. 1998; 37:407–419. [PubMed: 9704982]

Balleine BW, Killcross AS, Dickinson A. The effect of lesions of the basolateral amygdala on instrumental conditioning. J Neurosci. 2003; 23:666–675. [PubMed: 12533626]

Balleine BW, O'Doherty JP. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology. 2010; 35:48–69. [PubMed: 19776734]

Colwill RM, Rescorla RA. Postconditioning devaluation of a reinforcer affects instrumental responding. Journal of experimental psychology. Animal behavior processes. 1985; 11:120–132.

Coutureau E, Killcross S. Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav Brain Res. 2003; 146:167–174. [PubMed: 14643469]

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. Nature. 2006; 441:876–879. [PubMed: 16778890]

De Diego-Balaguer R, Couette M, Dolbeau G, Durr A, Youssov K, Bachoud-Levi AC. Striatal degeneration impairs language learning: evidence from Huntington's disease. Brain. 2008; 131:2870–2881. [PubMed: 18842608]

Dickinson A. Actions and habits: the development of behavioural autonomy. Philosophical Transactions of the Royal Society of London. 1985; B308:67–78.

Dickinson A, Nicholas DJ, Adams CD. The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. Quarterly Journal of Experimental Psychology. 1983; 35B: 35–35 I.

Finch DM, Gigg J, Tan AM, Kosoyan OP. Neurophysiology and neuropharmacology of projections from entorhinal cortex to striatum in the rat. Brain Res. 1995; 670:233–247. [PubMed: 7538025]

Haber SN. The primate basal ganglia: parallel and integrative networks. J Chem Neuroanat. 2003; 26:317–330. [PubMed: 14729134]

Herrnstein RJ. Relative and absolute strength of response as a function of frequency of reinforcement. J Exp Anal Behav. 1961; 4:267–272. [PubMed: 13713775]

Hilario MR, Costa RM. High on habits. Front Neurosci. 2008; 2:208–217. [PubMed: 19225594]

Hilario MRF, Clouse E, Yin HH, Costa RM. Endocannabinoid signaling is critical for habit formation. Frontiers in Integrative Neuroscience. 2007; 1:6.10.3389/neuro.07/006.2007 [PubMed: 18958234]

Jin X, Costa RM. Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature. 2010; 466:457–462. [PubMed: 20651684]

Miyachi S, Hikosaka O, Lu X. Differential activation of monkey striatal neurons in the early and late stages of procedural learning. Exp Brain Res. 2002; 146:122–126. [PubMed: 12192586]

Miyachi S, Hikosaka O, Miyashita K, Karadi Z, Rand MK. Differential roles of monkey striatum in learning of sequential hand movement. Exp Brain Res. 1997; 115:1–5. [PubMed: 9224828]

Packard MG, McGaugh JL. Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: further evidence for multiple memory systems. Behav Neurosci. 1992; 106:439–446. [PubMed: 1616610]

Poldrack RA, Packard MG. Competition among multiple memory systems: converging evidence from animal and human brain studies. Neuropsychologia. 2003; 41:245–251. [PubMed: 12457750]

Seger CA. How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. Neurosci Biobehav Rev. 2008; 32:265–278. [PubMed: 17919725]

Seger CA, Cincotta CM. The roles of the caudate nucleus in human classification learning. J Neurosci. 2005; 25:2941–2951. [PubMed: 15772354]

Solomon M, Frank MJ, Smith AC, Ly S, Carter CS. Transitive inference in adults with autism spectrum disorders. Cogn Affect Behav Neurosci. 2011; 11:437–449. [PubMed: 21656344]

Squire LR. Memory systems of the brain: a brief history and current perspective. Neurobiol Learn Mem. 2004; 82:171–177. [PubMed: 15464402]

Staddon JE, Cerutti DT. Operant conditioning. Annu Rev Psychol. 2003; 54:115–144. [PubMed: 12415075]

Ullman MT. A neurocognitive perspective on language: the declarative/procedural model. Nat Rev Neurosci. 2001; 2:717–726. [PubMed: 11584309]

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]

Wiltgen BJ, Silva AJ. Memory for context becomes less specific with time. Learn Mem. 2007; 14:313–317. [PubMed: 17522020]

Wiltgen BJ, Zhou M, Cai Y, Balaji J, Karlsson MG, Parivash SN, Li W, Silva AJ. The hippocampus plays a selective role in the retrieval of detailed contextual memories. Curr Biol. 2010; 20:1336–1344. [PubMed: 20637623]

Yin HH. The sensorimotor striatum is necessary for serial order learning. J Neurosci. 2010; 30:14719–14723. [PubMed: 21048130]

Yin HH, Knowlton BJ. Contributions of striatal subregions to place and response learning. Learn Mem. 2004; 11:459–463. [PubMed: 15286184]

Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci. 2004; 19:181–189. [PubMed: 14750976]

Yin HH, Knowlton BJ, Balleine BW. Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur J Neurosci. 2005a; 22:505–512. [PubMed: 16045503]

Yin HH, Knowlton BJ, Balleine BW. Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav Brain Res. 2006; 166:189–196. [PubMed: 16153716]

Yin HH, Mulcare SP, Hilario MR, Clouse E, Holloway T, Davis MI, Hansson AC, Lovinger DM, Costa RM. Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nat Neurosci. 2009; 12:333–341. [PubMed: 19198605]

Yin HH, Ostlund SB, Knowlton BJ, Balleine BW. The role of the dorsomedial striatum in instrumental conditioning. Eur J Neurosci. 2005b; 22:513–523. [PubMed: 16045504]

Zanini S, Tavano A, Fabbro F. Spontaneous language production in bilingual Parkinson's disease patients: Evidence of greater phonological, morphological and syntactic impairments in native language. Brain Lang. 2010; 113:84–89. [PubMed: 20188407]

A) RATE OF LEVER PRESSING DURING TRAINING

B) RATE OF HEAD ENTRIES DURING TRAINING

C) DEVALUATION TEST

Figure 1. Different dorsal striatal circuits are involved in goal-directed actions and habits in mice

(**A**) Acquisition of lever pressing in DLS and DMS sham and lesioned animals trained under ratio and interval schedules of reinforcement. (**B**) Average rate of head entry throughout training in DLS and DMS sham and lesioned mice. (**C**) Devaluation Test: lever pressing during the valued versus the devalued condition for the different training schedules in DLS and DMS sham and NMDA-lesioned animals, normalized individually to the lever pressing of the last day of training.
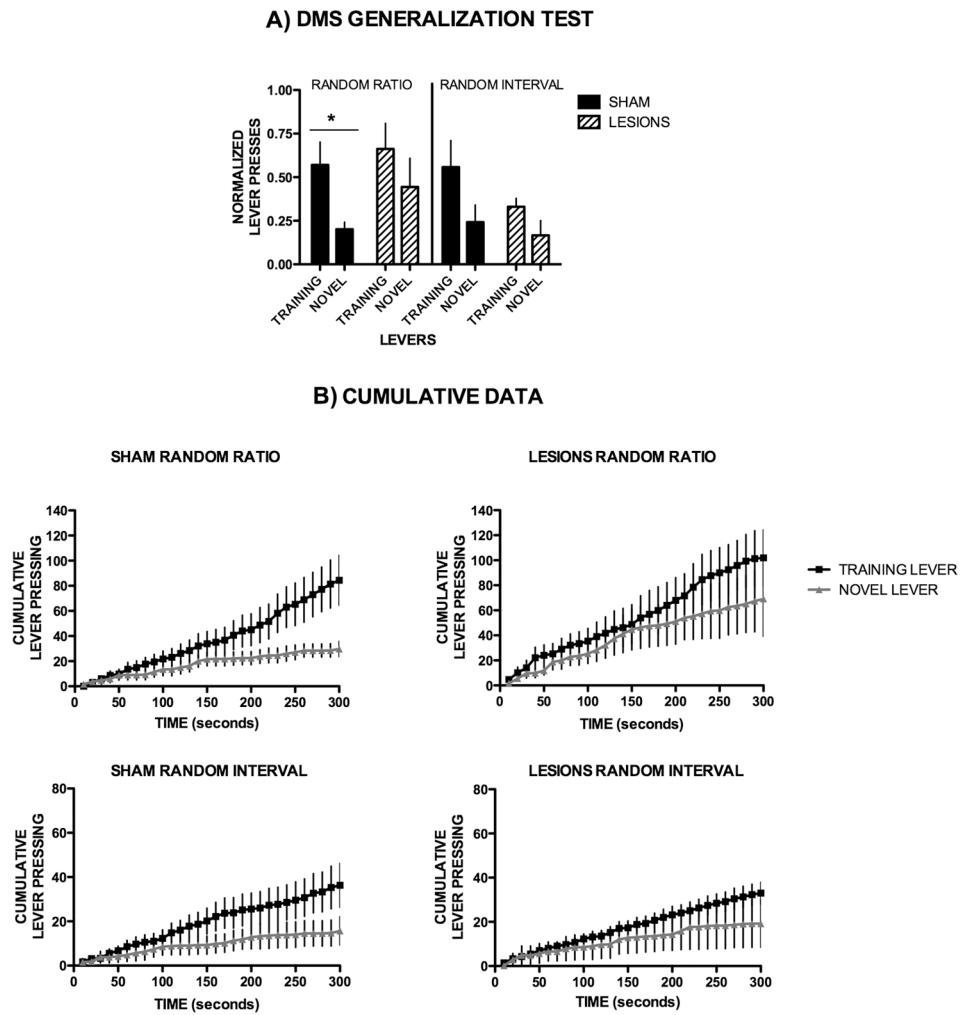
A) DLS GENERALIZATION TEST
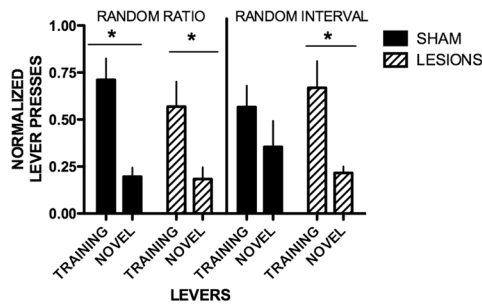

B) CUMULATIVE DATA
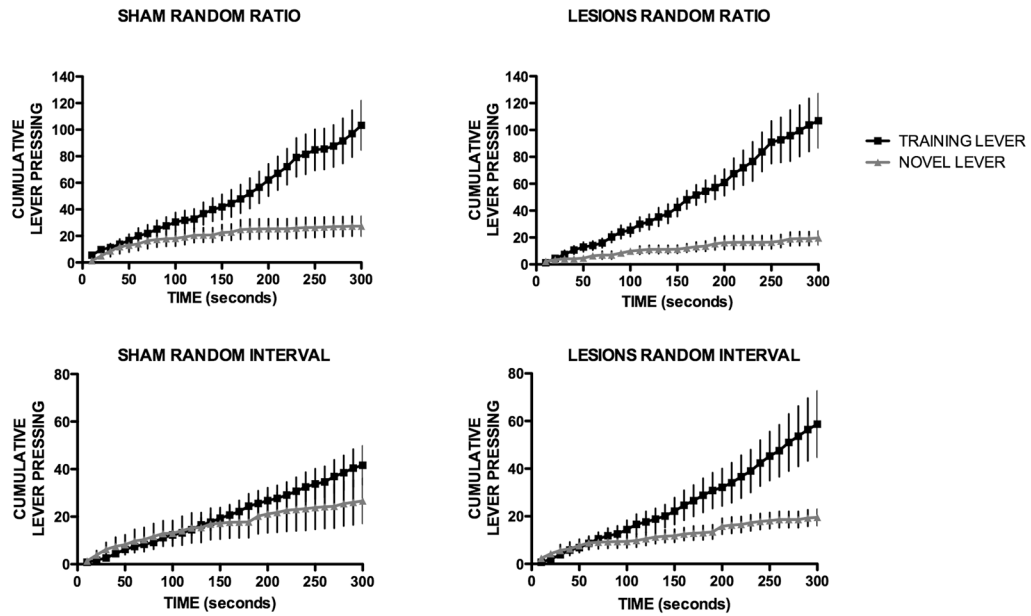
Figure 2. Dorsolateral striatum is critical for action generalization
(**A**) Lever pressing on the training lever versus a novel lever for the different training schedules of DLS sham and lesioned mice, normalized individually to the lever pressing of the last day of training. (**B**) Cumulative lever pressing for the training lever vs. the novel lever during the generalization test for DLS sham and lesioned mice.

A) DMS GENERALIZATION TEST

B) CUMULATIVE DATA

Figure 3. Dorsomedial striatum is necessary for action discrimination
(A) Lever pressing on the training lever versus a novel lever for the different training schedules of DMS sham and lesioned mice, normalized individually to the lever pressing of the last day of training. **(B)** Cumulative lever pressing for the training lever vs. the novel lever during the generalization test for DMS sham and lesioned mice.
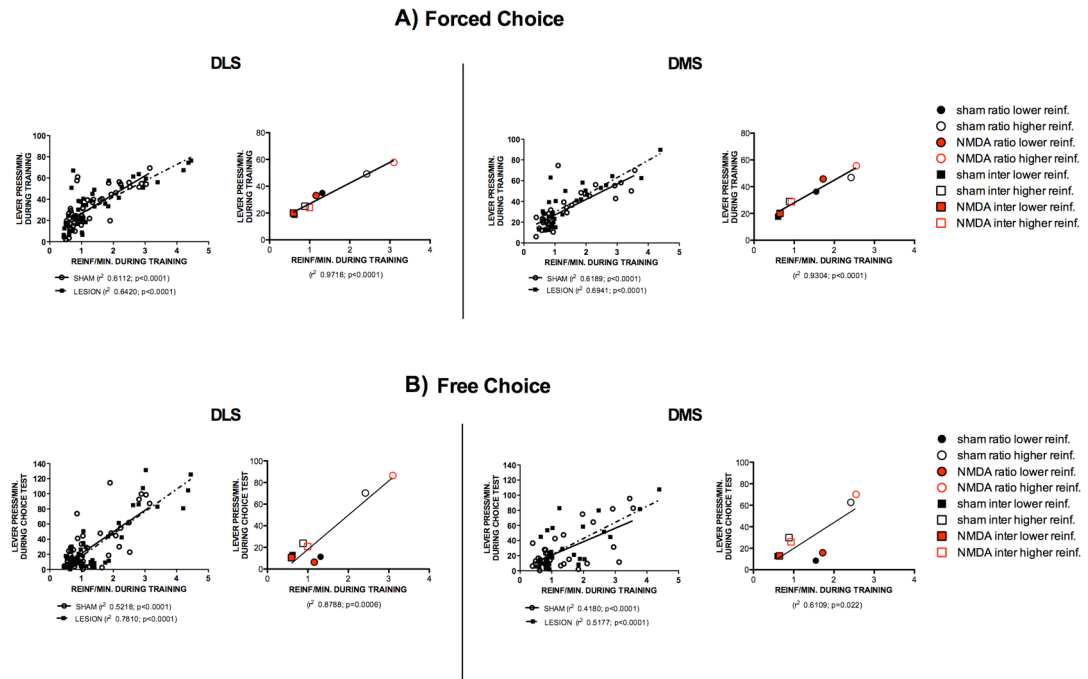
Figure 4. Dorsal striatum lesions do not affect the ability of mice to match/explore
(A) Correlations between lever pressing rates and reinforcement rates in DLS and DMS lesioned and sham animals during forced-choice training. Left panels: Correlations between the average lever presses for each animal during the 4 days of forced-choice training and reinforcement rate during the same period, for lesioned and sham animals. Right panels: Correlations between the average lever presses for each group of animals during the 4 days of forced-choice training and reinforcement rate during the same period **(B)** Correlations between lever pressing rates during the free-choice test and reinforcement rates during training. Left panels: Correlations between the lever presses of each animal during the free-choice test and reinforcement rate during training, for lesioned and sham animals. Right panels: Correlations between the lever presses for each group of animals during the free-choice test and the reinforcement rate during training.