NEURAL BASIS OF PROXIMAL INFLUENCES ON CHOICE: RECENT REINFORCEMENT, INTER-TRIAL INTERVAL, AND UNCERTAINTY

RAJAT THAPA
Bachelor of Science, University of Lethbridge, 2012

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

Department of Neuroscience
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

NEURAL BASIS OF PROXIMAL INFLUENCES ON CHOICE:
RECENT REINFORCEMENT, INTER-TRIAL INTERVAL, AND UNCERTAINTY

RAJAT THAPA

Date of Defence: May 3, 2018

Dr. A. Gruber Supervisor	Associate Professor	Ph.D.
Dr. R. Sutherland Thesis Examination Committee Member	Professor	Ph.D.
Dr. P. Hayes Thesis Examination Committee Member	Professor	Ph.D.
Dr. G. Metz Internal Examiner	Professor	Ph.D.
Dr. A. Greenshaw External Examiner University of Alberta Edmonton, Alberta	Professor	Ph.D.
Dr. A. Luczak Chair, Thesis Examination Committee	Associate Professor	Ph.D.

Dedication

For mom.

Abstract

Describing how animals make decisions is fundamental to understanding animal behaviour. Here we describe the transient yet significant influence of recent wins and losses on subsequent decisions in a competitive two-choice paradigm. We report that the drive to shift to alternate choices after a loss (lose-shift response) decays monotonically within the first few seconds of the inter-trial interval whereas the tendency to repeat a choice after a win (win-stay response) is initially low, gradually increases for few seconds, and then declines. We also show that the level of contextual uncertainty is correlated with the number of exploratory responses observed in the operant chamber. Furthermore, we report that lose-shift is dependent on the integrity of the ventrolateral striatum whereas win-stay is dependent on proper phasic dopaminergic signalling in the ventromedial striatum. Overall, our results suggest that lose-shift and win-stay response depend on dissociable circuits within the ventral striatum.

Acknowledgements

Graduate school has been the best time of my life. Thanks to the generosity of the lengthy list of people that helped me along the way, I have had the unreal opportunity of being part of the scientific process of discovery under the tutelage of world-class neuroscientists. Supervising graduate students is probably like raising toddlers—unformed piece of clay needing much moulding and remoulding. Therefore, I would like to foremost thank my supervisor Dr. Aaron J. Gruber for not only tolerating me for the past five years but going above and beyond the call of duty countless times in shaping proper thinking habits in me. To quote Aaron in one instance: “Too many people fall into the trap of over-interpreting their data or inventing narratives…let’s just shoot for revealing the ground truth and present the data with as little bias as possible.” This allegiance to honesty and data-driven thinking has served me well throughout my graduate career and will continue to serve me as the recipe for progress for rest of my life.

I would like to especially thank Dr. Robert J. Sutherland for being the mentor that I could always look up to. He pursues research not only out of humble curiosity but with a sense of duty. His efforts attracted the NSERC CREATE grant to Canadian Centre for Behavioural Neuroscience (CCBN) which funded this Ph.D. project among others. I would like to equally thank Dr. Paul G. Hayes and Dr. Robert J. McDonald of my supervisory committee for always making time for the meetings and providing valuable suggestions for progressing the project. I am grateful for all my teachers that made neuroscience study exciting: Dr. Matthew Tata for forever changing the way I see sound, Dr. Artur Luczak for invaluable

statistics and MATLAB skills, Dr. Andrew Iwaniuk for anatomy lessons, Dr. David Euston for discussion on prefrontal cortex, Dr. Robbin Gibb for foundational neuroscience methods training, Dr. Gerlinde Metz for encouraging fearless questioning, and Dr. Sergio Pellis for setting the foundation for understanding behaviour. I would also like to thank Dr. Ian Whishaw, Dr. Bruce McNaughton, Dr. Peter Shizgal of Concordia University, and Dr. Kent Berridge of the University of Michigan for being both an incredible mentor and an inspiration.

Special thanks must be given to all my friends and colleagues at CCBN who made work enjoyable. Among them I would like to particularly mention Scott Wong, Victorita Ivan, Saeedeh Hashemniayetorshizi, Cecilia Badenhorst, Clifford Donovan, Ali Mashhoori, Sienna Randolph, Leo Molina, Maurice Needham, Ivan Skelin, Afrooz Jalali, Josh Heynen, Logan Persson, Alicia Briggs, Aubrey Demchuk, Valerie Lapointe, Melinda Wang, Khadaryna Hernandez, James Cardiff, Tine Gulbrandsen, Simon Spanswick, Fraser Sparks, Keiko McCreary, Mashal Fida, Kartik Iyer, Shweta Soni, Surjeet Singh, Hendrik Steenland, Moira Holley, Karen Dow-Cazal, Isabelle Gauthier, Doug Bray, Naomi Cramer, Marg McKeen, Amanda Mauthe-Kaddoura and Roberto.

None of this would be possible without the unending love, understanding, and support provided by my mom and sisters. Last but not the least, Kriszty and Ale, you are my reason for anything I do.

Thank you.

Table of Contents

List of Tables

List of Figures

showing a decrease within sessions but no effect of barrier length. (L, inset) Plots of lose-shift and licking for each barrier condition, showing that licking is not sufficient to account for variance in lose-shift between barrier conditions. Statistically significant difference among group means: *p < 0.05, *** p < 0.001. Error bars show SEM.

Figure 4. Effect of consecutive wins or losses on choice: test for reinforcement learning. A, Plot of probability of a stay response on trial n, after a win (i.e., win-stay; left) or win-stay-win sequence (right) for each rat. The latter is the probability that the rat will chose the same feeder in three consecutive trials given wins on the first two of the set. The data show an increased probability of repeating the choice given two previous wins on the same feeder compared with a win on the previous trial, consistent with RL theory. B, Plot of probability of a switch response on trial n after a loss (lose-shift; left) or after a lose-stay-lose sequence (right). The probability of shifting after two consecutive losses to the same feeder is not greater than the probability of shifting after a loss on the previous trial, which is inconsistent with the predictions of RL theory. In both plots, gray lines indicate a within-subject increase in probability, whereas red lines indicate a decrease. ***Statistical significance of increased probability (p < 0.001) within subjects.

Figure 5. Responses during every training session for one cohort. Responses plotted for each rat (symbol-color) and each day of training. Session 1 is the second time the rats were placed into the behavioural box, and reward probability was p = 0.5 for each feeder regardless of previous responses or rewards. A, Number of trials completed in each session. Rats were allowed 90 min to complete up to 150 trials in sessions 1–10, and hallways of increasing lengths were introduced in sessions 3–8. B–D, Plot of the probability of responding to the rightward feeder, probability of lose-shift, and probability of win-stay during the first 16 sessions. The majority of rats showed no side bias, strong lose-shift, and very little win-stay in initial trials. Only a few rats showed initial side bias, and therefore little lose-shift and strong win-stay (blue shading in panels B–D). Lose-shift was invariant over training, whereas win-stay increased (see text). Dark lines indicate median across all subjects for each day.

Figure 6. Task apparatus and responding. (A) Schematic representation of the behavioural apparatus and examples of operant sequences on the task. Valid sequences consist of a nose poke in the poke port followed by locomotion to one of the two feeders. Rats sometimes chose to locomote from one feeder to the other without committing a nose poke; we term this extraneous feeder sampling (EFS). (B) The probability of EFS immediately after reward (win) or reward omission (loss) for each rat (Cohort 1: n = 68 for this and subsequent panels), showing that reinforcement does not affect EFS likelihood. (C) The probability of lose-shift responding following trials with EFS (green) or no EFS (black) parsed into bins of inter-trial-interval. EFS dramatically reduces lose-shift probability regardless of ITI for the population. (D) The within-subject plot of mean lose-shift probability. (E) Mean lose-shift probability for each rat computed from either the first feeder chosen after the nose poke or the last feeder chosen before the subsequent nose poke.

Nearly all rats appeared to generate lose-shift responses from the last feeder chosen as compared to the first feeder chosen, suggesting that the EFS strongly influences subsequent choice. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from the comparison group ($p < 0.000001$).

Figure 7. Changes in EFS and win-stay-lose-shift responding within and between sessions. (A) The mean probability of EFS decreases over the training sessions (Cohort 1: n = 68). (B, C) The mean probability of lose-shift or win-stay responding does not change over the training sessions. (D-F) Correlations among the probability of EFS, lose-shift, and win-stay responding among rats on the last day of training. The immediate effect of EFS on win-stay and lose-shift measures were minimized by omitting trials following EFS. EFS was uncorrelated with the other response types. (G-H) The plot of the probability of EFS, lose-shift, and win-stay (dashed line) for bins of 30 trials within sessions. Only EFS decreased within sessions. (I) The plot of EFS probability versus trial bin (10 trials/bin) within each of several sessions of a separate cohort of rats (Cohort 2, n = 30), showing that within-session variance of EFS reduces with training. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from the comparison group ($p < 0.000001$).

Figure 8. Effect of devaluation on task performance. (A) Mean cumulative sum of operant responses (nose-poke to feeder) in bins of time within a session (Cohort 2: n = 30 rats in panels A-C). Pre-feeding rats 20 minutes before the task reduced the number of trials performed. (B) The mean cumulative sum of EFS events in the same sessions, which was not reduced by pre-feeding. (C) The mean relative rate of EFS/trials for each pre-feeding level, showing an increase with devaluation. (D-F) Same plots as above for a new heterogeneous cohort collected by different experimenters (Cohort 3: n = 48 in panels D-F), showing replication of the devaluation effects. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from the comparison group ($p < 0.003$).

Figure 9. Effect of mid-session change in the barrier on EFS rate. Mean relative EFS rate within a session before and after the barrier was replaced at trial 101 (male subjects from Cohort 3: n =16). Replacing the barrier with either a longer (red dashed line) or shorter (blue dotted line) barrier increased EFS as compared to replacing with the same length barrier (black solid line). Asterisks (*) indicate a significant difference of means by post-hoc analysis ($P < 0.04$).

Figure 10. Effects of lesions of the dorsolateral striatum (DLS) or nucleus accumbens core (NACc). (A-B) The extent of the excitotoxic striatal lesions in Cohort 4 (n = 21). The black and grey shading show minimal and maximal extent of the lesions to the dorsolateral striatum (DLS, n = 7) or nucleus accumbens core (NACc, n = 7), respectively. (C) Mean response times, showing that DLS-lesioned rats made slower responses than the other two groups (n = 7 controls). (D) The mean percentage of rewarded trials was not affected by lesions. (E) Mean number of licks prior to reward. (F) Cumulative sum of completed task trials in bins of time within sessions. (G) Cumulative sum of EFS, showing no reduction in lesioned rats

relative to controls. (H) Relative rate of EFS to operant responses (trials) within sessions, showing a dramatic increase in DLS-lesioned animals. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from each other by Tukey's HSD post-hoc test ($p < 0.01$).

Figure 11. Effect of striatal lesions on lose-shift and win-stay responding. (A) The group-averaged probability of lose-shift responding, showing that DLS-lesioned animals decreased lose-shift relative to the other groups, and approached the optimal level in this task ($p = 0.5$). (B) The plot of the probability of lose-shift versus the logarithm of the inter-trial-interval for each group. The DLS-lesioned animals show low lose-shift regardless of the ITI. (C) The group-averaged probability of win-stay responding. (D) The plot of the probability of win-stay versus the logarithm of the inter-trial-interval, showing that animals with NACc lesions have reduced win-stay probabilities regardless of ITI. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from each other by Tukey's HSD post-hoc test ($p < 0.05$).

Figure 12. Behavioural task and histology. (A) Schematic representation of the competitive choice task apparatus and the operant response. (B) A representative Nissl-stained brain section along with a schematic representation of the largest (light grey) and smallest (dark shading) electrolytic lesions of the lateral habenula. (C) flowchart of the sequence of events and the definition of lose-shift, win-stay, and extraneous feeder approach (EFS) in the competitive choice task.

Figure 13. Effects of LHb lesions on win-stay and lose-shift responses in the competitive choice task. (A) Probability of lose-shift responding by the two experimental groups on the test day. (B) Correlation between the probability of lose-shift responding and the stereological estimates of the percentage of LHb damaged in each subject. Animals plotted at 0% are controls. $r = 0.056$. (C) Probability of lose-shift responding as a function of inter-trial-interval (ITI). The black dotted lines represent data from the control group, whereas the grey dotted lines represent the LHb lesion group. (D) Probability of win-stay responding. (E) Correlation between the probability of win-stay responding and the stereological estimates of the percentage of LHb damaged in each subject. $r = -0.582$. (F) Probability of lose-shift as a function of ITI. Error bars represent 95% confidence intervals, and asterisks (*) indicate statistically different means ($p < 0.05$).

Figure 14. Effects of LHb lesions on locomotion in the competitive choice task. (A) Number of trials completed within a session in 5-min time bins. (B) Ratio of extraneous feeder sampling (EFS) to the number of operant trials. (C) Number of 'patch visits', which is the sum of the number of distinct entries made to the feeders and nose-port. (D) Average time taken to reach a feeder after making a nose-poke. (E) Total number of licks made at either of the feeder wells. (F) Percentage of trials that were rewarded with sucrose. Error bars represent 95% confidence intervals, and asterisks (*) indicate statistically different means ($p < 0.05$).

List of Abbreviations

ACC Anterior Cingulate Cortex

AMPH D-amphetamine

DLS Dorsolateral Striatum

DMS Dorsomedial Striatum

EFS Extraneous Feeder Sampling

lOFC lateral Orbitofrontal Cortex

ITI Inter-trial Interval

LHb Lateral Habenula

mOFC medial Orbitofrontal Cortex

NACc Nucleus Accumbens core

OFC Orbitofrontal Cortex

PrL Prelimbic Cortex

VLS Ventrolateral Striatum

VMS Ventromedial Striatum

Chapter one

General Introduction

Animals learn, adapt, and make up their own patterns of behaviour. This behavioural flexibility is the topic of this thesis. Specifically, I discuss the proximal determinants of adaptive decisions. To understand our approach in investigating such choices, consider the example of someone flipping a coin for a thousand times. The coin comes up either heads or tails in any given flip and for the full experiment, there is a different mixture of heads and tails for the series of tosses. The relative frequency of heads and tails tells us something fundamental about the nature of coin flipping. The outcome is equiprobable, but the experimenter is unsatisfied with the odds and wants to improve the predictability of the toss. He rigs up a mechanical coin flip machine that can apply and measure exact forces on each flip and additionally records the time the coin is in the air. Soon he starts to discover that although on average the probability of a flip is 50%, on any given trial he can predict the outcome with significantly greater accuracy when he knows the force applied, the time the coin was in the air, whether the head or tail was facing up before the toss, the height the coin reaches, and other ‘proximal’ variables. The experiments outlined in this thesis are similar in nature to the second experiment. In trying to explain how our model organism (the rat) decides between two options in a series of choices, we investigate the trial-to-trial proximal variables that have a significant impact on the rat’s next choice. We find that indeed a brain trying to guide behaviour on a trial-by-trial basis not only integrates the sum of previous experiences but also incorporates decision heuristics and proximal perceptual evidence.

Explanations of animal behaviour range from ultimate causes such as pressures driving natural selection to proximate mechanisms of an animal's physiology, for instance, hormones or neurotransmitter function (Tinbergen 1963). Here we limit our discussion to the learning and memory mechanisms that shape behaviour and guide choices. Among the various influences on choice, rewards and punishment have long maintained a disproportionate attention from behavioural scientists given its power to predict and shape behaviour. In 1911 Edward Lee Thorndike promulgated the law of effect (Thorndike 1911). In brief, the law of effect predicts that rewards strengthen the bond between an animal's response that leads to the reward and the stimulus setting in which the response is made and increases the probability of repeating the response. In contrast, punishment or reward omissions lead to a weakening of that bond and consequently decreases the likelihood of repeating the same response in the future (Thorndike 1927). The ensuing behaviourists discovered that an animal's choices are influenced not only by the outcomes of the choices but also by when and how often the rewards are delivered (Ferster and Skinner 1957). Skinner and his peers advocated two distinct levels of behavioural analyses of the effects of reinforcement: moment-by-moment influences on the shaping of new behaviour and the strengthening of a choice over time. However, in the 1960's when Skinner's Harvard pigeon lab was passed on to R. J. Herrnstein, he emphasized the quantitative possibilities inherent in Skinner's second kind of analyses. He described the relations among response rate, probability, and strength and formulated it into the ‘matching law’ which states that when animals make choices between two alternatives, the ratio of response rates to the two choices equals the ratio of reinforcements yielded by each response (Herrnstein 1961). For example, if choice A was rewarded 60% of the time and choice B was rewarded 40% of the time, the animals making recurrent choices can match their allocation of choices close to the probability of reward in each choice. Herrnstein’s analyses of behaviour had such an influence on the scientific community that it shifted the discussion of the effects of reinforcements heavily towards investigations of behavioural patterns over time and away from moment-by-moment analysis of behavioural variability (Shimp 2014). And even today proximal moment-by-moment influences on choice are largely ignored in favour of average behavioural patterns over time.

However, trial-by-trial behavioural variations sometimes clearly affect the temporal behavioural pattern under investigation. For instance, in experiments where animals are making recurrent choices between two options, the matching law can break down due to undermatching when subjects too often shift between the two responses (Baum 1974). To avoid undermatching, experimenters typically introduce a forced inter-trial interval termed changeover delay (COD) lasting few seconds in the testing procedure to stop rapid switching between alternatives that may be caused by the influence of local contingencies of reinforcement on choice. With COD in place, animals easily acquire a win-stay response strategy—if a choice leads to reward, animals choose the same choice in the consecutive trial.

B. A. Williams (1991), however, observed that local contingencies had a larger influence on choice with shorter COD. And, with longer COD, the choice proportion was determined more reliably by the matching law. What are these local influences on behaviour that have power over choice above and beyond the law of effect? In this thesis, we investigate these largely ignored proximal variables that have a considerable effect on behavioural variability.

All the experiments in this thesis used the competitive choice task (CCT). The CCT apparatus and procedure are described in detail in Chapter two. In brief, behavioural testing was performed in operant boxes containing two liquid delivery feeder wells and a central nose-poke port that were separated by a barrier orthogonal to the wall (Figure. 1A). Individual trials of the task began with the illumination of two cue lights mounted proximally to the nose-poke port and the inactivation of the overhead house light. Animals then had 15 seconds to commit a nose-poke into the central port, and subsequently locomote to one of the two possible feeder wells of which only one was baited in each trial. If the rat selected the rewarded feeder, it received a drop of 10% sucrose solution (Win). If the rat chose the non-rewarded well, the feeder was left empty (Lose) and the house light was illuminated. The computer selected a priori which feeder well to reward based on a ‘competitive’ algorithm that uses the well choice and reinforcement history of the rat to predict the choice in the current trial and minimize the number of rewarded trials (Lee et al. 2004). The optimal solution for the rat in the competitive choice task is to be as random as possible by distributing its choices equally between the two feeder wells.

Several features of CCT distinguishes it from other operant tasks. Most operant conditioning experiments employ one of two types of schedule of reinforcement. An animal can be put on a continuous reinforcement schedule where every correct response is rewarded. On the first day of training in CCT, we employed such a continuous reinforcement schedule to facilitate learning. However, most operant tasks employ a partial reinforcement schedule where the animal is rewarded only in some trials because partial reinforcement is more motivating than continuous reinforcement (Ross 1964). Some partial reinforcement experiments employ either a fixed or a variable ratio schedule where the animal is rewarded after a certain number of responses (e.g. 10 lever presses), whereas other experiments use fixed or variable interval schedules where the reward is doled out after a certain amount of time has passed. CCT does not fall under either of these categories because the task is self-paced and has components of both variable ratio and variable interval schedule of reinforcement. CCT is different from Pavlovian tasks as well. There are no explicit cues informing the animal of the correct response. Furthermore, the operant response is composed of two distinct serial responses: first, to enter the centre lane to make a nose-poke response and second, to enter one of the two reward-well lanes to check for sucrose delivery. The animal needs to have an intact serial order memory to be able to make correct responses. These unique combinations of features of the CCT place it in a novel category of operant tasks.

In this thesis, we focus on three behavioural responses in the CCT. First, we investigate the lose-shift response which is the tendency of animals to switch to the alternative after a reward omission (loss). We report that the lose-shift tendency varies significantly between trials based on the inter-trial interval (ITI). The probability of lose-shift is highest at the shortest ITI and decreases rapidly to chance level (0.50) within about seven seconds (Chapter two). Second, we examine the probability of win-stay behaviour which is the tendency of animals to repeat a choice given it was rewarded. We report that win-stay has an inverted U-shaped relationship with ITI i.e., the probability of win-stay steadily increases until about 8 seconds and then decreases (Chapter two). Third, we investigate a response we termed extraneous feeder sampling (EFS) which is direct shuttling behaviour between the two feeder wells. EFS was never rewarded and indeed delayed access to the next reward. Nonetheless, animals continued to produce a significant number of EFS responses even after extended training. We provide evidence for the theory that the frequency of EFS is correlated with contextual uncertainty in Chapter three. Together, the behavioural measures lose-shift, win-stay, and EFS examine the proximal influence of recent wins and losses, inter-trial interval, and uncertainty on choice.

In Chapter three, four, and five we explore the neural basis of lose-shift and win-stay behaviour. Traditionally, learning from wins and losses has been attributed to reward prediction error-based dopaminergic signals (Glimcher 2011). However, we report that while win-stay depends on the proper dopaminergic input to the ventromedial striatum, lose-shift is not affected by manipulations of dopamine levels either systemically or locally in the striatum. Earlier studies from our lab implicated the lateral striatum in lose-shift responses (Skelin et al. 2014).

The lateral striatum, however, is composed of functionally and anatomically distinct subregions: dorsolateral striatum and ventrolateral striatum. Chapter five presents the case that focal damage of the ventrolateral striatum is sufficient to attenuate the lose-shift response and lesions of dorsolateral striatum most likely disrupt serial order memory necessary for operant responses in the CCT.

The win-stay and lose-shift decision strategy could in principle derive from a reinforcement learning mechanism with a large learning rate (Sutton and Barto 1998). However, the evidence presented in Chapters two, three, four, and five consistently support our proposal that lose-shift and win-stay depend on dissociable brain circuits. In brief, win-stay and lose-shift have vastly different temporal relationship with inter-trial interval (Chapter two; Gruber and Thapa 2016); they are differentially affected by dopamine level manipulation (Wong et al. 2016), and lesions of striatal regions produce differential effects on win-stay and lose-shift such that lesions of ventromedial striatum affect win-stay but not lose-shift whereas lesions of ventrolateral striatum eliminate lose-shift but not win-stay. In summary, results presented in this thesis support the theory that the learning and memory systems subserving the influence of proximal wins and losses on decision-making are not only dissociated from classical reinforcement learning mechanisms but also from each other.

Chapter two

Lose-shift Responding Decays Rapidly After Reward Omission and is Distinct from Other Learning Mechanisms in Rats[1]

Abstract

The propensity of animals to shift choices immediately after unexpectedly poor reinforcement outcomes is a pervasive strategy across species and tasks. We report here that the memory supporting such lose-shift responding in rats rapidly decays during the inter-trial interval and persists throughout training and testing on a binary choice task, despite being a suboptimal strategy. Lose-shift responding is not positively correlated with the prevalence and temporal dependence of win-stay responding, and it is inconsistent with predictions of reinforcement learning on the task. These data provide further evidence that win-stay and lose-shift are mediated by dissociable neural mechanisms and indicate that lose-shift responding presents a potential confound for the study of choice in the many operant choice tasks with short inter-trial intervals. We propose that this immediate lose-shift responding is an intrinsic feature of the brain's choice mechanisms that is engaged as a choice reflex and works in parallel with reinforcement learning and other control mechanisms to guide action selection.

[1] Chapter published as: Gruber, A. J., & Thapa, R. (2016). The memory trace supporting lose-shift responding decays rapidly after reward omission and is distinct from other learning mechanisms in rats. eneuro, 3, ENEURO. 0167-0116.2016. The thesis chapter contains revisions different from the published article. Student's Contribution: Data collection and analysis, literature review, and proofreading manuscript.

Introduction

Animals use various strategies when choosing among responses yielding uncertain reinforcement outcomes. These strategies may be informed by the discounted sum of many past reinforcements so as to bias choice toward actions that, on average, have provided more favourable reinforcements (Herrnstein 1961). This is embodied by reinforcement learning and other algorithms that can account for experience-dependent choice bias that evolves over many trials (Rescorla and Wagner 1972; Sutton and Barto 1998). The evidence that mammals can use a form of reinforcement learning to solve some tasks is overwhelming (Balleine and O'Doherty 2010). The brain, however, has several other robust systems for learning and memory that can strongly influence actions (McDonald and White 1995; Stote and Fanselow 2004; Gruber and McDonald 2012). For instance, decisions are often disproportionately influenced by the recency of reinforcements and other proximal factors, which are not fully captured by conventional reinforcement learning algorithms; such factors can be included via additional model components so as to improve the fit of these algorithms to behavioural data from rodents and humans (Ito and Doya 2009; Rutledge et al. 2009; Skelin et al. 2014). In particular, animals performing operant tasks for appetitive outcomes tend to repeat responses that were rewarded in immediately preceding trials (win-stay), whereas they tend to shift to alternative choices if preceding responses were not rewarded (lose-shift). These choice strategies have been reported in many studies spanning a wide array of tasks and species, including humans (Frank et al. 2007; Wang et al. 2014), nonhuman primates (Mishkin et al. 1962; Schusterman 1962; Lee et al. 2004), rats (Evenden and

Robbins 1984; Skelin et al. 2014), mice (Means and Fernandez 1992; Amodeo et al. 2012), pigeons (Rayburn-Reeves et al. 2013), and honeybees (Komischke et al. 2002). It is important to identify the neural mechanisms of these ubiquitous strategies to improve neurobiologically grounded theories of choice behaviour.

Lose-shift responding in an operant task may be abolished by lesions of the sensorimotor striatum in rats (Skelin et al. 2014), which is unexpected because this striatal region has been predominantly associated with the gradual formation of habits that are relatively insensitive to changes in reinforcement value as revealed by devaluation procedures and maze navigation (Yin et al. 2004; Pennartz et al. 2009). Here we reveal several new dissociated properties of lose-shift and win-stay response strategies, which can account for some apparent discrepancies in findings from distinct testing paradigms. In particular, we show that the temporal dependence of choice on previous reinforcement can present a significant confound pertinent to an array of behavioural tests and can account for the involvement of the dorsolateral striatum in the present task but not in devaluation.

Methods

Subjects

A total of 115 male Long-Evans rats (Charles River, Saint-Constant, QC, Canada, except as noted below) were used in the experiments presented here. The correlation and temporal dependence of lose-shift and win-stay responding were determined from all animals that met the performance criterion (n = 98; see below); these data were collected over 14 months by four different experimenters in five distinct cohorts. The subject information (number of subjects, age at time of first testing) are as follows: n = 17, 88 d; n = 19, 110 d; n = 16, 100 d; n = 46, 99 d;

and n = 17, 126 d. One of these cohorts (n = 19) was additionally used for the barrier experiment described below (Figure 3). A different cohort (n = 17), which was born in-house, was used to study task acquisition (Figure 5). The protocol for the behavioural task and the testing apparatus used was the same for each experiment. All animals weighed 350–600 g at the time of testing and were pair-housed in standard clear plastic cages in a vivarium with a 12-h light/dark cycle (lights off at 7:30 p.m.). The animals were allowed to habituate in the facility and were handled for at least 2 min/d for 1 week before training. Behavioural training and testing were conducted during the light phase (between 8:30 a.m. and 6:00 p.m.). The animals were restricted to 1 h of water access per day in individual cages and had ad libitum access to water on weekends; body weight was maintained at > 85% of pretesting weight. All animal procedures were performed in accordance with the authors' university animal care committee's regulations.

Apparatus

Behavioural training and testing took place in one of six identical custom-built aluminum boxes (26 X 26 cm). Each box contained two-panel lights and two liquid delivery feeders on either side of a central nose-poke port (Figure 1A). Infrared emitters and sensors in the feeders and central port detected animal entry. After illumination of the panel lights, a rat poked its snout into the central port to initiate a trial, and then responded by locomoting to one of the two feeders. Each feeder was equipped with an optical beam break system in the feeder to detect licking. The beam was conducted to the indentation in the feeder where the liquid reward was delivered via a pair of plastic optical fibers, and rapid changes in transmitted light intensity were detected with an industrial red/infrared (680-nm) emitter/sensor unit designed for detecting rapid interruptions in transmission while self-adjusting the emitted light power to counteract slow changes (Banner Engineering, Minneapolis, MN, model D12DAB6FP). This system sometimes registered the entry of liquid into the feeder and could sometimes count a break as two events (the on and the off phases) because of the self-adjusting feature. The number of detected licks may, therefore, be biased. These biases are invariant over time, and we randomized the assignment of subjects to testing boxes each session; the biases, therefore, do not present significant obstacles for interpreting the relative changes in licking behaviour. A 13-cm-long aluminum barrier orthogonal to the wall separated each feeder from the central port. This added a choice cost and reduced choice bias originating from body orientation. A longer 20-cm barrier was used in some sessions to increase the inter-trial interval (ITI) between reward feeder exit and the subsequent nose-poke to begin the next trial. Control of the behavioural task was automated with a microcontroller (Arduino Mega) receiving commands via serial communication from custom software on a host computer. The hardware connections from the microcontroller to the sensors, valves, and lights were made via optically isolated solid-state relays (Crydom, IDC5, ODC5). We attempted to reduce acoustic startle from sounds outside of the testing chamber by presenting constant background audio stimuli (local radio station).

Behavioural task

Individual trials of the task began with illumination of two panel-mounted lights mounted proximally to the nose-poke port and inactivation of the overhead house light. Animals then had 15 s to commit a nose-poke into the central port and subsequently respond to one of the two possible feeders. If the rat failed to respond, the apparatus would briefly enter an error state, in which the house light would illuminate, and the panel lights would extinguish. The state of the lights was then reset (house light off; panel lights on) with a delay of 100–500 ms, which depended on the communication latency of the microcontroller with the host computer. The computer selected a priori which feeder to reward. If the rat selected this rewarded feeder, it received a 60-µl drop of 10% sucrose solution (a win) with a short delay determined by the hardware (typically < 50 ms) and the fluid dynamics of the solution in the delivery system. The state of the lights did not change. If the rat chose the nonrewarded feeder, it was left empty (lose) and the apparatus would switch to the error state (house light on; panel lights off) for 100–500 ms until the system reset. This brief change in lighting was intended to signal that reward was not forthcoming. This delay was shorter than the time required for rats to locomote from the feeder to the nose-poke port when barriers are present, and so did not implement a timeout penalty. We, therefore, consider the task to be self-paced within the 15-s limit on trial duration. The computer-implemented a "competitive" algorithm similar to previous studies (Lee et al. 2004; Skelin et al. 2014). Briefly, the algorithm examined the entire choice and reward history in the present session to exploit predictable responses and minimize the number of rewards delivered. This was done by using sequences of the most recent (the previous four) choices and reinforcements as a pattern to determine the probability of choosing each feeder based on the entire previous choice history within the current session. If the algorithm detected that either feeder was chosen more than chance in this context from all previous trials in the current session (probability > 0.5 by the binomial test,

$p < 0.05$), it was selected to be unrewarded for that trial. The competitive algorithm, therefore, punished predictable response patterns. The optimal solution for the rat was to be as stochastic as possible in feeder choice. Daily sessions of the task were 45 min in duration, and rats were randomly assigned a starting time and testing box for each session.

All animals were trained on the competitive choice task by gradually introducing components of the task. Initially, there were no barriers between the central port and feeders, and 50% of responses were rewarded. Subsequent sessions used the competitive algorithm. The barrier separating the nose-poke port and feeders was increased in discrete lengths (4, 8, and 13 cm) over several sessions (typically four to five). Training was complete when animals performed > 150 trials with the 13-cm barrier within the 45-min session over two consecutive days. Training terminated for any subject that had not met the criterion by at least 2 d after 50% of the other members of the cohort had met the criterion. Animals typically completed training on sessions 8–11. Note that the termination of training and the inclusion criterion of 100 trial/session were implemented in an attempt to homogenize experience across cohorts. Although some rats are slow to acquire the task, less than 5% fail to acquire the task with additional training. The training and inclusion criteria do not likely bias the subjects strongly toward select phenotypes. We modified the training schedule for one cohort (noted in Results) by limiting trials to 150 trials per day so that acquisition across sessions would be more homogeneous among subjects.

Analysis

Data included up to two sessions per rat; sessions were included only if rats performed at least 100 trials (n = 98 rats). Population means were computed from means for each subject computed across all sessions (one point from each subject). Data were analyzed and plotted with custom-written code and built-in function of Matlab 2015a (Mathworks, Natick, MA), with the exception of ANOVA using a within-subjects design (a.k.a. RM-ANOVA), which was conducted with IBM SPSS V21 (IBM Canada, Markham, ON, Canada). We report the number of trials computed as the sum total number of complete trials within a session. We limited analyses related to reward dependence to trials in which the rat sampled only one reward feeder between trials. This was done to eliminate any effect of visiting the second feeder (the one not initially chosen) before the next trial. The probability of lose-shift was calculated as the probability that the subject would shift feeder choice in trials after reward omission. Likewise, the probability of win-stay was calculated as the probability that the subject would repeat the selection of feeders on trials immediately after rewarded trials. In defining consecutive trials, we include only trials that were < 20 s apart.

We used the Matlab function “fitnlm” for fitting function parameters to the relationship between response switching probability and ITIs. Fits were weighted by the number of samples used to compute probabilities to minimize the effect of variance in the data points derived from low numbers of samples. We quantitatively validated the model fits to data collected under different conditions by computing the difference in the area under the curve (AUC) for each animal in each condition. We chose constant-sized bins in linear time that had sufficient samples in all conditions (n > 25) to compute probabilities. Invariance of conditions should thus result in no change in the AUC. Some subjects did not have sufficient samples in each ITI bin for each condition (e.g., non-overlapping ITI distributions due to motoric slowing) and were excluded from the AUC analysis. We also used the models of the ITI–probability relationships to estimate the motoric effects of treatments, which otherwise present confounds of the treatments on response probabilities. To do this, we used the Matlab function “predict” to predict the expected change in the probability of lose-shift for each animal according to its change in the base 10 logarithms of ITIs across conditions (barrier) using the model.

We compared the predictive power of a standard reinforcement learning algorithm called Q-learning (Watkins and Dayan 1992) with that of win-stay/lose-shift. We first fitted the parameters of the Q-learning algorithm independently to best fit each animal’s responses on one session as described previously (Skelin et al. 2014). We then used these parameters on the same data to compute the predicted most likely next response for each subject. Note that we are using the same data for testing and training, which yields the best possible accuracy of the model. The predictions for the win-stay/lose-shift algorithm were simply determined by the reinforcement and choice on the previous trial using the same sessions as for the Q-learning model and required no parameter fitting. We then computed the prediction accuracy for each model as the percentage of correct predictions.

The power of statistical tests was computed with SPSS for ANOVA or the software package G*Power (http://www.gpower.hhu.de/en.html) for other analyses (see Table 1). Superscript letters listed with p-values correspond to the statistical tests shown in Table 1.

Results

Lose-shift and win-stay responding are uncorrelated and have distinct time dependences

We trained Long Evans rats to perform a non-cued binary choice task in which they entered a nose-poke port to initiate a trial and then locomoted to one of two liquid sucrose feeders on either side of a barrier (Figure 1A). A computer algorithm computed which feeder was to be rewarded on each trial and attempted to minimize the number of rewards delivered by first using the reward and choice history of the rat to predict its next feeder choice, and then selecting the alternate feeder to be rewarded (Lee et al. 2004; Skelin et al. 2014). The optimal strategy is a random choice on each trial; win-stay, lose-shift, or other predictable response are suboptimal on this task and result in a rate of reinforcement less than the expected maximum of 50%. Deviation from a random strategy reveals features about the brain's learning, memory, and choice mechanisms. We examined 44,898 trials from 98 rats run in five different cohorts over 14 months. The rats in our sample performed well on the task; they collected reward on a mean of 46.0 ± 4.3% of trials (range: 41.7 - 55.3%) compared with an expected maximum of 50%. This is in line with the performance of nonhuman primates (47 - 48%) and rats (42 ± 1.4%; Tervo et al., 2014) on similar tasks.

We next examined how delivery (win) or omission (lose) of reward affected rats' choice on the subsequent trial of the task. The population showed very robust lose-shift responding (68.8 ± 1.0% of trials; t-test that mean is 50%: t(97) = 19.2, p = 1E–34[a]), but not win-stay responding (51.6 ± 11.9% of trials; t-test, t(97) = 1.4, p = 0.17[b]). These strategies were negatively correlated among subjects ($r^2 < 0.25$, F(97) = 32.2, p = 1E–6[c]; Figure 1B). In other words, nearly every subject showed lose-shift responding, and the more likely they were to shift after losses, the less likely they were to stay after wins. We next investigated how the effect of reinforcement on subsequent choice depends on time. Trials of the task were self-paced, and we computed ITIs as the time between the first exit of the reward feeder and the next entry into the poke port. This is the minimum amount of time that reward information, or its effect on choice, needs to be represented to affect the subsequent response. ITIs were longer after win trials than after loss trials (Figure 1C, E), which is qualitatively consistent with post-reinforcement pauses (Felton and Lyon 1966) and the frustrative effects of reward omission (Amsel 1958) long observed in other tasks in which animals receive a reward on only a fraction of responses. The temporal effects here are much shorter than past studies, and other reported pauses seem to depend on prospective motoric requirements rather than past actions (Derenne and Flannery 2007); it is, therefore, difficult to compare this aspect of our data to previous studies that have largely omitted the type of barriers we have used. As we show later, the longer ITIs after rewarded trials very likely involves the time spent licking and consuming the reward.

The effect of the reinforcement type (win/lose) on subsequent choice has a distinct dependence on ITIs. The probability of lose-shift responding has a prominent log-linear relationship with the ITI at the population level (r^2 = 0.96, df = 14; F statistic vs. constant model = 389, p = 1E–11[d]; Figure 1D), suggesting exponential decay of the influence of reward omission on subsequent choice in linear time. The probability of lose-shift reaches chance level (p = 0.5) within 7 s for the population. This could arise either because of a within-animal process (e.g., decaying influence) or because of individual differences among the population (e.g., faster rats have a stronger tendency to shift). To distinguish among these possibilities, we tested whether the relationship between ITI and lose-shift was evident within subjects. Indeed, this negatively sloped log-linear relationship does fit the behaviour of most individual subjects (t-test that slope of fit for individual subjects was equal to 0: t(54) = 40.0, p = 1E–40[e]; Figure 1D inset). This indicates that the temporal dependence occurs within individual subjects rather than exclusively at a population level.

In contrast to the log-linear temporal dependence of lose-shift, the probability of win-stay shows a log-parabolic relationship with ITI, in which it first increases to a peak at ~8 s before decreasing (r^2 = 0.60, df = 14, F statistic vs. constant model = 12.8, p = 1E–3[f]; Figure 1F). This log-parabolic relationship also fit the behaviour of most individual subjects (t-test that the distribution of quadratic coefficients fit to each subject has a mean of 0: t(63) = 6.6, p = 1E–8[g]; Figure 1F inset), again indicating a within-subject effect. In sum, the choices of most individual subjects in our large sample show dependence on the time since the last reinforcement, consistent with a temporally evolving neural process. Moreover, the distinct temporal profiles of lose-shift and win-stay responding support the hypothesis that they are mediated by distinct neural processes. The temporal dependencies of these response types have inverse slopes near the mean ITI after wins or losses, which thereby suggests an explanation for the negative correlation between the probability of win-stay and lose-shift. If the ITI after wins and losses are correlated within animals, then faster animals will show strong lose-shift and weak win-stay, whereas slower animals will have weaker lose-shift and stronger win-stay. Indeed, the ITI after wins is highly correlated with the ITI after losses (r^2 = 0.70, F(97) = 225, p =1E–26[h]), and the correlation between subject-wise mean lose-shift and the logarithm of the mean ITI after loss is moderately strong (r^2 = 0.17, F(97) = 20.6, p = 2E–5[i]). However, the correlation between mean log ITI after wins and win-stay among subjects is weak (r^2 = 0.03, F(97) = 1.8, p = 0.18[j]), likely because of the nonlinear dependence of win-stay on log ITI (Figure 1F) and because of between-subject variance in the acquisition of win-stay as described later. In sum, the inverse relationship between lose-shift and win-stay responding among subjects (Figure 1B) can likely be attributed to subject-wise variation in ITI.

We next tested how motivation may affect the prevalence of lose-shift responding by quantifying the variation of dependent variables within sessions. We computed means of variables over bins of 15 consecutive trials for each animal before generating population statistics. We presume that reasonable behavioural correlates of motivation in this task are the response time (from poke-port to feeder) and the number of licks made before reinforcement time (either reward delivery or panel lights extinguishing). As rats accumulate rewards within the session, we presume their motivation decreases, and thereby expect increased response time and decreased anticipatory licking. Indeed, response time increases after the first 15 trials (RM-ANOVA main effect trial: F(9,864) = 2.8, p = 0.003[k]; Figure 2A), whereas anticipatory licking decreases (RM-ANOVA: F(9,864) = 8.8, p = 1E–6[l]; Figure 2B). Furthermore, anticipatory licking correlates very strongly with the total number of licks on each trial (r^2 = 0.83, F(8) = 38.7, p = 3E–4[m]; Figure 2B inset), further supporting the notion that this metric reflects motivation. The decrease of licking within session contrasts the increase of lose-shift responding within sessions (RM-ANOVA: F(9,864) = 2.2, p = 0.02[n]; Figure 2C). Indeed, these are strongly, and negatively, correlated (r^2 = 0.78, F(8) = 27.8, p = 7E–4[o]; Figure 2C inset). This suggests that lose-shift responding is not driven by motivation. On the other hand, the ITI after losses decreases as sessions progress (RM-ANOVA: F(9,864) = 29, p = 1E–6[p]; Figure 2D), and this decrease (in log space) is correlated with increased lose-shift (r^2 =0.76, F(8) = 24.8, p = 1E–3[q]; Figure 2D inset). In sum, the movement speed to the feeders and anticipatory licking decrease within sessions; these changes likely reflect decreasing motivation during the session. On the other hand, the ITI after losses decreases, likely because in part of the reduced time spent licking in the feeders. Thus, the fact that lose-shift responding increases as sessions progress suggest it is more likely directly related to changes in ITI than is motivation, in agreement with the overwhelmingly strong correlational evidence of this relationship at the population and individual levels (Figure 1D). Of course, motivation almost certainly plays a role in modulating the ITI, and can thereby exert indirect effects on lose-shift responding.

Change in lose-shift responding is predicted by change in ITI: evidence for a decaying influence

The regression analysis of individual subjects’ responses indicates that the decrease in lose-shift responding that occurs as ITIs get longer is observed within most subjects. We hypothesize that this could reflect a decaying influence, analogous to decay or accumulating interference of short-term memory of other information (Mizumori et al. 1987; Altmann and Gray 2002). The previous correlation analysis is not sufficient to rule out alternative hypotheses, such as a population component to the phenomenon. For instance, rats with short ITIs may be more sensitive to reinforcement omission than rats moving more slowly. We thus tested these hypotheses by assessing whether the dependence of choice on previous reinforcement is altered by inducing longer ITIs. The choice–ITI curve should translate (shift) to the right with increased median ITI if the choice–ITI relationship is due to a population effect, but should remain invariant to increasing ITIs if the relationship is due to a decay of a decaying influence. We assessed this by alternating between short (13 cm) and long (20 cm) barriers on successive days for one cohort (n = 19 rats, six sessions). Rats presumably have the similar motivation (i.e., thirst) regardless of the barrier length. This is supported by the fact that the decrease in the number of trials (24.0%) is proportional to the increase in median ITI (25.6%) in the session with longer barriers, suggesting that the decrease in trials is due to increased locomotion time rather than decreased motivation to complete trials. Furthermore, the running velocity during responses (nose-poke to feeder) is not affected by the barrier length (19.0 cm/s for shorter and 19.1 cm/s for longer barrier; paired t-test of different means: t(18) = 0.05, p = 0.96[r]). Last, the amount of anticipatory licking in feeders is not affected by barrier length (reported below). Although the longer barrier increased ITIs after either losses or wins, neither the lose-shift or win-stay relationship with ITI was shifted by this procedure (Figure 3A-d). We tested this in two ways: first, qualitatively by computing the coefficient of determination (r^2) for population data from both the short and long barrier sessions with respect to the one common model fit for all data; and second, quantitatively by computing the difference in the area under the curve (AUC) for each subject across the two barrier conditions. The ITI bins and integration range are held fixed for the AUC computation in each barrier condition for each rat; translation or deformation of the ITI–probability curves induced by the barrier will, therefore, lead to different integration values, and the difference in AUC between barrier lengths will be nonzero. We computed the difference of AUC for each rat and tested for a nonzero population mean as a test for an effect of the treatment. For lose-shift, population data from each session fit the common model well (r^2_{short} =0.82; r^2_{long} = 0.68, df = 18), and there was no change in the mean difference of the AUC (t-test that mean difference is 0: t(16) = 0.09, p = 0.93[s]; Figure 3B inset). Likewise for win-stay, session population data from each condition fit the common model well (r^2_{short} = 0.69, r^2_{long} = 0.60, df = 18), and the mean difference of area under the curve across subjects was not different from zero (t(14) = 0.55, p = 0.59[t]; Figure 3D inset). Because the curves are invariant to increases in ITIs, these data support the hypothesis that the response phenomenon is a result of a within-subject factor such as a decaying influence rather than a population effect of motivation or movement speed.

The lose-shift probability decreases in sessions with longer barriers (paired t-test: t(18) = 4.7, p = 2E–4[u]; Figure 3E), and this change is accurately predicted by the increase in ITI (Figure 3F) using the log-linear model for each animal (t-test that mean change in lose-shift is the same as the model prediction: t(18) = 0.14, p = 0.89[v]; Figure 3G). In other words, the model can predict the change in lose-shift based on the change in median ITI for each rat. The overall probability of win-stay does not change (Figure 3H), which is expected because the change in ITI after wins is small with respect to the curvature of the win-stay relationship with ITI. The percentage of rewarded trials is higher in the sessions with the long barriers, as is expected because responses are less predictable when lose-shift responding decreases toward chance level (paired t-test that mean lose-shift is not increased: t(18) = 2.45, p = 0.02[w]; Figure 3I). Thus, the log-linear model accounts for several features of responding, providing strong evidence that it is an appropriate representation of the relationship between lose-shift responding and ITI.

We next investigated whether the barrier length affects within-session correlations, to additionally assess whether the changes could be due to changes in motivation or outcome valuation. The prevalence of lose-shift responding did not vary within the session, partly because of the high probability of lose-shift in the first few trials in the long barrier condition (main within-subject effect of trial RM-ANOVA: F(6,109) = 1.6, p = 0.16[x]; Figure 3J). The general trend, however, appears to be increasing lose-shift responding as the session progresses, consistent with the analysis in the previous section (Figure 2). Also consistent with this previous analysis, the post-loss ITI decreases (RM-ANOVA: F(6,109) = 5.7, p 3E–5[y]; Figure 3K), and anticipatory licking decreases (RM-ANOVA: F(6,108) = 6.8, p = 4E–6[z]; Figure 3L) as sessions progress. Note that we used all trials in the computation of anticipatory licking to increase samples, whereas we exclude trials after sampling of both feeders for the other metrics (see Methods). The longer barriers evoked a reduction of lose-shift responding (main within-subject effect of length RM-ANOVA: F(1,18) = 8.3, p = 0.01[aa]) and increase in ITI (RM-ANOVA: F(1,18) = 28, p = 5E–5[ab]) but evoked no change in licking (RM-ANOVA: F(1,18) = 0.5, p = 0.52[ac]) across the session. These data thus support our prediction that motivation decreases within sessions, and that the increased within-session lose-shift prevalence is driven by decreases in ITI after losses. Moreover, lose-shift is again moderately correlated with anticipatory licking in the shorter hallway condition (r^2 = 0.67; F(5)

= 10.1, p = 0.02[ad]; Figure 3L inset), but data collected in the long barrier condition do not fall on the same line. This indicates that some other factor (e.g., ITI) is needed to predict the relationship between them. This is in stark contrast to the single log-linear relationship between ITI and lose-shift that accounts for data from both barrier conditions (Figure 3B).

In sum, the data in this section provide very strong evidence that lose-shift responding decreases with increased barrier length not because the underlying mechanism changes, but rather because the distribution of the ITI shifts to the right (larger values) so that the decaying influence has more time to decay. This indicates that the form of the memory mechanism underlying lose-shift responding is invariant to the animals' movement speed, and the model can be used to predict changes in lose-shift responding based on changes in ITI.

Lose-shift responding in the task is inconsistent with reinforcement learning

We have previously shown that the addition of explicit terms for lose-shift and win-stay to a standard reinforcement learning (RL) model improves the prediction of rat choice behaviour on this task (Skelin et al., 2014). Moreover, RL does not provide a normative account of the rapid decay of lose-shift responding. Nonetheless, RL mechanisms may contribute to win-stay or lose-shift responding. For instance, a large learning rate will cause choice to be highly sensitive to the previous trial by driving large increases (decreases) of the choice after wins (losses). We thus tested a fundamental prediction of RL: successive wins or successive losses on the same choice should have an additive (albeit sublinear) effect on choice. For instance, the probability of a stay response after a win-stay-win sequence on the same feeder should be greater than that after a win irrespective of outcomes in the past. Formally, this is expressed by the inequality: Prob (stay$_n$ I win$_{n-1}$, stay$_{n-1}$, win$_{n-2}$) > Prob (stay$_n$ | win$_{n-1}$). Indeed, we find that the probability of staying after a win-stay-win sequence is greater than the probability of staying after a win (paired t test that the above equality is not true for rats with at least 25 samples of win-stay-win sequences: $t(48) = 10.2$, $p = 1E–13$[ae]; Figure 4A). Likewise, the probability of switching should be increased after a lose-stay-lose sequence, formalized by Prob (shift$_n$ | lose$_{n-1}$, stay$_{n-1}$, lose$_{n-2}$) > Prob (shift$_n$ | lose$_{n-1}$). However, we find that this is not the case (paired t test that the above equality is not true for all rats with at least 25 samples of lose-stay-lose sequences: $t(32) = 2.2$; $p= 0.99$[af]; Figure 4B). Thus, the probability of shifting is not increased after two consecutive losses at one feeder compared with the probability of shifting after loss on the previous trial, which is inconsistent with the foundational concept of RL that the value of the feeder should be additionally decremented by the second loss, and therefore the likelihood of choosing the other feeder should be higher (e.g., shift). In sum, the RL concept of reinforcement-driven value learning is consistent with responding after wins, but not after losses. This suggests that the neural mechanisms involved in lose-shift are distinct from those involved in RL. Conventional RL has the facility to implement win-stay-lose-shift, although not to the extent evident in the present data. To evaluate the predictive power of a standard RL algorithm (Q-learning) compared with a pure win-stay/lose-shift strategy, we computed the prediction accuracy for each model on one session from each rat in a cohort (see Methods, n = 19). The win-stay/lose-shift correctly predicted 60 ± 1% of responses, whereas Q-learning predicted 52 ± 1% of responses. It is worth noting that we tested the prediction of Q-learning on the same data that were used to fit the model parameters so as to produce the highest possible accuracy regardless of overfitting. Nonetheless, these data provide strong evidence that the win-stay/lose-shift strategy better accounts for responding on this task (t-test of mean prediction accuracy between models: $t(34) = 5.2$, $p = 1E–5^{ag}$). It is not surprising that RL does not account for responses on this particular task because the expected long-term utility of both feeders is equivalent. If the probability or amount of reward were unequal at the two feeders, the brain would likely engage RL systems to overshadow the lose-shift mechanisms presenting here.

Lose-shift responding is stationary during training, whereas win-stay is acquired

We next sought to determine whether the prevalence of lose-shift responding is related to aspects of the task, such as the competitive algorithm or barriers, which are atypical of other tasks. We, therefore, examined the probability of lose-shift and win-stay in a new cohort of rats (n = 17) undergoing a modified training schedule. In an attempt to normalize learning across subjects, rats were allowed 90 min in the behavioural box to perform up to a maximum of 150 trials per session over the first 10 days, and then unlimited trials for 90 min in subsequent sessions (Figure 5A). Increasingly longer barriers were introduced in sessions 3–8. A few rats initially had a strong side bias (blue shaded region in Figure 5B), and consequently tended to stay regardless of loses or wins (blue shaded region in Figure 5C, D). The majority of rats, on the other hand, showed prominent lose-shift responding across all sessions, even during the second session in the apparatus in which the competitive algorithm was not used and the probability of reward was p = 0.5 regardless of previous choices. Nonetheless, the probability of lose-shift (median = 0.86, including the animals with side bias) was significantly higher than chance on this session (two-sided Wilcoxon signed-rank test for median = 0.5, n = 17, p = 0.03[ah]). Moreover, the probability of lose-shift in the population did not vary across the training sessions (within-subjects main effect of session RM-ANOVA: F(15,150) = 0.54, p = 0.91[ai]; Figure 5C). In contrast, the probability of win-stay was initially less than chance (Wilcoxon; n = 17, p = 0.01[aj]) and increased across testing sessions (RM-ANOVA: F(15,150) = 2.3, p = 5E–3[ak]; Figure 5D). These features of the data would be even stronger by omitting the sessions in which rats had a large side bias. These data reveal that lose-shift responding is prevalent across all sessions, whereas win-stay is acquired during training, again supporting the hypotheses that they are mediated by separate processes. The pronounced lose-shift responding in the first several sessions indicates that it is not the barriers or competitive algorithm that induces animals to utilize this response strategy.

Discussion

The present data show that reward omission has a pronounced short-lasting effect on subsequent choice, which can be described by the classic notion of lose-shift responding that decays over several seconds. Several features of rodent lose-shift are distinct from those of win-stay: (a) their probability is not positively correlated among subjects; (b) their temporal dependence on ITI is dissimilar; and (c) lose-shift is prevalent from the first day of training and does not diminish, whereas win-stay is acquired during training. These data provide further evidence that win-stay and lose-shift are mediated by dissociated neural mechanisms. The temporal dependence of lose-shift responding presents a confound for the study of choice in rodents and other animals that likely influences performance in the many operant choice tasks with short ITIs. Moreover, manipulations that affect ITI (e.g., drugs, stress) will alter the prevalence of lose-shift and win-stay responses. Studies that do not control for this are difficult to interpret because tasks solvable by lose-shift will be facilitated by reduced ITI independently of other putative mechanisms. Lose-shift responding is thus an important latent variable to consider in behavioural studies of choice.

The highly prominent lose-shift responding over the 7-s interval considered here is not explained by conventional RL theory. In particular, the rapid decay of switching probability during the ITI has no normative basis in RL. Furthermore, that the probability of shifting after a lose-stay-lose sequence is not greater than that of shifting after a single loss is counter to the fundamental prediction of RL that subsequent losses on the same feeder should decrement the value of the action and therefore increase the probability of switching. On the other hand, the properties of win-stay are more consistent with RL, in that consecutive wins do increase the probability of a stay response. The dependence of win-stay on ITI, however, remains unexpected. This dissociation is counter to conventional RL formulations, in which wins and losses influence choice by modulating a singular value attached to actions or outcome states (Watkins and Dayan 1992; Sutton and Barto 1998). We instead propose that the lose-shift phenomenon can be characterized as an intrinsic choice reflex because of its prevalence in the task (despite being a non-optimal solution), its failure to diminish over thousands of trials, its reliable time course, and its apparent independence of neural systems involved in executive functions (Skelin et al. 2014).

The brief lose-shift system involving the sensorimotor system studied here is dissociated from the reinforcement learning signals in ventral striatum and orbitofrontal cortex observed in many other studies (Samejima et al. 2005; Daw et al. 2006; Paton et al. 2006; Matsumoto et al. 2007; Schönberg et al. 2007; Hori et al. 2009; Ito and Doya 2009; Bromberg-Martin et al. 2010; Gan et al. 2010; Alexander and Brown 2011; Day et al. 2011). We argue that lose-shift is an adjunct to RL in the guidance of choice; the neural mechanisms for RL likely solve problems requiring processing of value or utility over many trials to establish responding rates to various choice options, whereas the lose-shift mechanism likely introduces exploration among the choice only on a trial-by-trial scale. The behavioural purpose of its 7-to-8 s time course is unclear, but this temporal window is supported by some of the few other reports that provide relevant evidence. Direct optogenetic activation of D2DR-expressing striatal cells in the dorsal striatum of mice results in place avoidance for about 10 s (Kravitz et al. 2012), and the behavioural effect of losses during a lever-pressing task is observed only when ITIs are less than ~15 s (Williams 1991). Moreover, win-stay/lose-shift behaviour is prominent in pigeons only when ITIs are ~10 s (Rayburn-Reeves et al. 2013). The emergence of rapidly decaying lose-shift behaviour across species and tasks, even when it is not needed or is suboptimal, suggests it is a general feature of choice intrinsic to its underlying mechanisms.

The decaying influence supporting lose-shift is only one of several memory systems in the brain. Rats can maintain information related to reinforcement over much longer intervals, and performance on these longer-interval tasks is often sensitive to disruption of the prefrontal cortex (Euston et al. 2012). This suggests that goal-directed behavioural control involving prefrontal cortex has a longer memory frame than the one considered here. We speculate that this difference in time frame accounts for the discrepancy of our results from that of devaluations experiments, which indicate that behaviour mediated by sensorimotor striatum is not sensitive to changes in the affective value of reinforcements (Yin et al. 2004; Quinn et al. 2013). This result has had a profound influence on many current theories of choice (Daw et al. 2005; Balleine and O'Doherty 2010; Gruber and McDonald 2012; van der Meer et al. 2012). In devaluation, the affective state of the animal is altered with either satiation or illness paired with the outcome, and the memory time is hours to days. Many regions of the prefrontal cortex and subcortical limbic structures encode affective information over time periods spanning minutes to months (Euston et al. 2012) and project heavily to the medial and ventral striatum (McGeorge and Faull 1989; Vertes 2004; Voorn et al. 2004). It is not surprising, then, that devaluation depends on medial striatum and not dorsolateral striatum. The lose-shift phenomenon studied here occurs over several seconds, and reward omission likely does not elicit a strong affective component because rats are denied only a small amount reward on each trial relative to the total reward intake over the session. We, therefore, propose that the inverse role of the dorsolateral striatum in devaluation and lose-shift behaviours derive from the differences in memory time interval and sensory domain (affective vs. sensory). In other words, the sensorimotor systems have explicit access to recent sensory information (including that related to reward) needed for lose-shift, but not direct access to remote affective information as needed for devaluation effects.

The percent of rewarded trials in our sample is on par with that of rats (Tervo et al. 2014) and nonhuman primates (Lee et al. 2004) competing against the same algorithm, albeit with different motoric demands. The probability of lose-shift alone is not reported in either study, but the non-human primates show only a slight amount of win-stay/lose-switch (Prob = 0.53 – 0.57) in the competitive task, similar to humans (Prob = 0.54 – 0.57; Hu et al. 2010). We speculate that the primate prefrontal cortex normally suppresses lose-shift by the sensorimotor striatum so that primates lose-shift less than rats. Rats appear to strongly use sensorimotor systems to respond during the task and therefore exhibit high amounts of lose-shift throughout training and testing.

Lose-shift responding is suboptimal in the present task, but its persistence is not likely to be an artifact of the task design. Lose-shift responding is prevalent on the second day of training without barriers and without the competitive algorithm and is invariant across training and testing. Other experiments have also revealed that rats do not perform optimally on binary choice tasks with dynamic reinforcements (Sul et al. 2011). Last, lose-shift is pervasive across many species and tasks (Mishkin et al. 1962; Schusterman 1962; Olton et al. 1978; Evenden and Robbins 1984; Means and Fernandez 1992; Komischke et al. 2002; Lee et al. 2004; Frank et al. 2007; Amodeo et al. 2012; Rayburn-Reeves et al. 2013; Skelin et al. 2014; Wang et al. 2014), and those abovementioned studies that report timing effects are consistent with the decay in our data. In sum, several lines of indirect evidence indicate that the lose-shift phenomenon studied here is not unique to the task, but rather appears to be a default strategy in many situations and is therefore relevant to many other behavioural tests with short ITIs.

The properties of lose-shift revealed here suggest it is an intrinsic feature of neural choice mechanisms in the striatum that can be described as a choice reflex; it is unlearned, prevalent in multiple cohorts, persistent, has a reliable time course, and involves the sensorimotor striatum. As such, the addition of explicit terms in RL models that include these properties will likely continue to improve model fits to data, particularly in tasks with short ITI and sensorimotor solutions (Ito and Doya 2009; Rutledge et al. 2009; Skelin et al. 2014).

In conclusion, lose-shift responding plays a simple but important role in trial-by-trial choice adaptation in some situations, particularly those with repetitious actions and rapid trials, and appears to work in parallel with reinforcement learning and other control mechanisms in dissociated neural structures to guide choice. Our data provide further evidence that theories of sensorimotor striatum function related to choice behaviour must expand from the current focus on gradual sensory-response associations and habit formation (Jog 1999; Daw et al. 2005; Balleine and O'Doherty 2010; Gruber and McDonald 2012; van der Meer et al. 2012) to also include rapid response adaptation that is dependent on a decaying influence.

Chapter three

Feeder Approach Between Trials is Increased by Uncertainty and Affects Subsequent Choices[2]

Abstract

Animals quickly learn to approach sources of food. Here, we report on a form of approach in which rats made volitional orofacial contact with inactive feeders between trials of a self-paced operant task. This extraneous feeder sampling (EFS) was never reinforced and therefore imposed an opportunity and effort cost. EFS decreased during initial training but persisted thereafter. The relative rate of EFS to operant responding increased with novel changes to the operant chamber, reward devaluation by prefeeding, or lesions to the dorsolateral striatum. We speculate that this may function to increase exploration when the task is uncertain (early in learning or introduction of novel apparatus components), when the opportunity cost is low, or when the learned sensorimotor solution is compromised. Moreover, EFS strongly affected subsequent choices by triggering a lose-shift response away from the sampled feeder, even though it occurred outside of the trial context. This indicates that at least some behaviours occurring between trials impact future behaviours and should be considered in decision-making studies.

[2] Chapter published as: Gruber, A. J., Thapa, R., & Randolph, S. H. (2017). Feeder approach between trials is increased by uncertainty and affects subsequent choices. eneuro, ENEURO. 0437-0417.2017. The thesis chapter contains revisions different from the published article. Student's Contribution: Performing surgeries, behavioural data collection and analysis, literature review, and proofreading manuscript.

Introduction

Optimal reward collection requires the ability to adjust behaviour based on past reinforcements and to inhibit unproductive actions (Thorndike 1927). In Reinforcement Learning Theory, the decision-maker’s level of knowledge about the task determines whether an action is productive or not (Sutton and Barto 1998). If there is no uncertainty because the decision-maker has full knowledge, then all directed actions should exploit the best source(s) of reward at a rate dictated by need, cost, and risk. Otherwise, the decision-maker should intersperse exploitative actions with some exploratory actions to gain information (Staddon and Motheral 1978; Kakade and Dayan 2002; Daw et al. 2006). Exploration allows for discovery of better reward sources or shortcuts to obtain known sources. In practice, humans and animals produce a variety of non-optimal actions in laboratory tasks (Chapter two; Breland and Breland 1961; Kahneman and Tversky 1979; Sugrue et al. 2004; Gruber and Thapa 2016). Although some can be attributed toward gaining information, much is attributed to a neurobiological failure to execute the optimal action policy or to inhibit underproductive (impulsive) actions (Moeller et al. 2001; Gruber et al. 2010; Bari and Robbins 2013).

Impulse control is a composite of processes that span motor, reward/effort, and choice domains (Evenden 1999; Aron 2011; Bari and Robbins 2013). Impulsive actions are often underproductive in laboratory tasks because they lead to suboptimal reward rates, either through smaller reward outcomes (Aparicio 2001; Reynolds et al. 2002), termination of trials (Carli et al. 1983), or because animals engage in actions that do not lead to reward (Breland and Breland 1961). Little attention has been given to the influence of such actions on subsequent behavioural choice (Evenden and Robbins 1984; Williams 1991). Here we investigate a form of unproductive behaviour that we refer to as extraneous feeder sampling (EFS); this occurs when animals ignore task contingencies and choose to make contact with feeders rather than begin the next trial (Figure 6). This is never reinforced and thus imposes an opportunity cost by consuming time and energy that could otherwise have been spent performing trials to collect the reward.

Animals often learn quickly to approach feeders, even when this is not required for reward delivery, as in the goal-tracking response in Pavlovian conditioned approach (Boakes 1977; Farwell and Ayres 1979; Robinson and Flagel 2009). Goal-tracking is reduced by outcome devaluation (Morrison et al. 2015), and the nucleus accumbens core is critical for the expression of Pavlovian conditioned approach (Parkinson et al. 1999; Blaiss and Janak 2009). We would expect comparable properties of our EFS phenomena if it involves a Pavlovian component. Moreover, Pavlovian-related learning and memory systems have long been proposed to influence instrumental actions and other behavioural output (Estes and Skinner 1941; Mowrer 1947; Rescorla and Solomon 1967). This likely arises from interactions among distinct behavioural control systems, which in some cases appear to function as opponent processes (Solomon and Corbit 1974; Boakes 1977). For instance, pigeons will peck at a stimulus (a Pavlovian-driven action) rather than collect reward via instrumental responding (Williams and Williams 1969). Moreover, rats approach and engage in operant responding on nearby levers more than distal ones, even if the nearby levers are associated with smaller rewards, require more effort, or impose longer delays to reward (du Hoffmann and Nicola 2014). This suggests that the brain systems involved in approach do not utilize information about relative outcome values, and it raises the important question of whether approach events can influence future actions, possibly by engaging learning in behavioural control systems that do represent outcomes.

Here, we sought to determine if EFS affects choice on subsequent trials and if EFS is related to task uncertainty, impulsivity, or Pavlovian control. Our data suggest that it is related primarily to uncertainty and can affect choices occurring many seconds later involving a different brain structure. This cross-talk between dissociated behavioural control systems is likely important for the study of choice in rodents and possibly other animals.

Methods

Subjects

This study involved 4 cohorts of Long-Evans (LE) rats (n = 170 total animals). Cohort 1 consisted of 68 male LE rats obtained from Charles River (Saint-Constant, QC, Canada) weighing between 450 and 600 g (postnatal day 94-102) at the time of behavioural testing. All rats are outbred wild-type unless noted otherwise. Cohort 2 consisted of 30 male LE rats (Charles River, Saint-Constant, QC, Canada) weighing between 350 and 450 g (postnatal day 88-106) at the beginning of behavioural testing. Cohort 3 consisted of 16 male and 6 female wild-type LE rats, and 14 male and 15 female LE rats expressing Cre-recombinase under the tyrosine hydroxylase (TH:Cre) born on site and weighing between 200 and 600 g (postnatal day 75-116) at the time of behavioural testing. Cohort 4 consisted of 21 male LE rats obtained from Charles River, Saint-Constant, QC, Canada and weighing between 450 and 600 g (postnatal day 94) at the time of behavioural testing. Housing conditions, training, and testing methods were common to animals from all cohorts. Rats were housed in pairs in a transparent plastic cage with corncob bedding and a section of PVC pipe for enrichment. Access to water was restricted to one hour per day during behavioural training and testing but was unrestricted otherwise. The vivarium was maintained at 21°C and 12-h light/dark cycle (lights off at 7:30 pm). Experimenters handled the rats daily for one week before the beginning of training. All experimental procedures were approved by the University of Lethbridge Animal Welfare Committee and adhere to the guidelines of the Canadian Council on Animal Care.

Competitive Choice Task

The competitive choice task (CCT) was used in all experiments. Behavioural training and testing took place in six identical custom-built aluminum boxes (26 X 26 cm). Each box contained two cue lights mounted proximally above the poke-port and two liquid delivery feeders on either side (Figure 6A). Infrared emitters and sensors in the feeders and central port detected animal entry. Following the illumination of the cue lights, the rats poked their snout into the central port to initiate a trial and then responded by locomoting to one of the two feeders. A 13-cm-long aluminum barrier orthogonal to the wall separated each feeder from the central port. This added a choice cost and reduced choice bias originating from body orientation. Control of the behavioural task was automated with a microcontroller (Arduino Mega, Italy) receiving commands via serial communication from custom software on a host computer. We reduced acoustic startle from sounds outside of the testing chamber by presenting constant background audio stimuli (local radio station).

All animals were trained on the CCT by gradually shaping components of the task. Initially, there were no barriers between the central port and feeders. Each trial of the task began with the illumination of the two cue lights. At this stage, the animals discovered that every poke port entry and a subsequent entry to either feeder within 15 s resulted in a reward of 60 µL of 10% sucrose solution. Once rats performed 150 trials (typically in the first session), the session was terminated. In the following session, feeder entry was rewarded with a probability of 0.5. Subsequent sessions used the competitive algorithm (described below). A barrier separating the nose-poke port and feeders was increased in discrete lengths (4, 8, and 13-cm) over several sessions (typically 4-5). The training was complete when the animals performed at least 150 trials with the 13-cm barrier within the 45-min session over two consecutive days (typically 7-10 training sessions in total).

A computer program served as a competitor for the rats and was implemented as in previous studies (Algorithm 2; Barraclough et al. 2004; Lee et al. 2004; Skelin et al. 2014; Gruber and Thapa 2016). The algorithm attempts to predict the rat’s next choice by comparing the pattern of choice sequences in the preceding trials (1-4 back) with the choice history of the current session. If any of the patterns occurred more likely than chance (computed by the binomial test), the algorithm baited the least likely feeder to be selected on the current trial. If no pattern was detected, the rewarded side was picked randomly. The optimal response policy of the rat is to choose randomly on each trial and disregard reinforcements. The statistical power of the algorithm to detect patterns is initially very weak, and so selects the rewarded feeder randomly for the first several trials.

Devaluation

Rats were trained on the CCT and divided into three groups. After all, subjects met the training criterion, individuals of each group received free access to a limited amount of the reward (sucrose solution) 20 minutes prior to the start of the CCT. The amount of pre-feeding was counterbalanced among rats so that an approximately equal number of rats received each of the three pre-feeding volumes (0, 5, 10 ml) each testing day. The volume given to each group rotated each of three consecutive days so that each rat had received one of the three levels prior to behavioural testing.

Excitotoxic lesions

Surgeries were performed after training was complete in a new group of rats (cohort 4). Rats were then randomly assigned to one of three lesion groups: dorsolateral striatum (DLS, n = 7); nucleus accumbens core (NACc, n = 7); or control (n = 7). All rats received Buprenorphine (Alstoe Ltd., UK) to mitigate pain 30 min before incision. The animals were anesthetized using 4% isoflurane gas (Benson Medical Industries Inc., Ontario, Canada) in oxygen flowing at 1.0 L/min and the surgical plane was maintained with 2% isoflurane throughout the surgery. The animals were mounted on a stereotaxic frame (Kopf Instruments, Tujunga, CA, USA). DV was measured at bregma and lambda to ensure the head surface was levelled. A midline incision was made to expose the skull. Burr holes were drilled through the skull to allow lowering of infusion cannulas at the following coordinates from bregma [in mm (AP, ML, DV)]: LS (1.6, 3.0, -6.2), (0.8, 3.7, -6.6), (-0.5, 4.5, -6.6); NACc (1.2, 2.1, -7.8). Bilateral lesions of LS and NACc were achieved by microinfusion of quinolinic acid (30 mg/mL in Dimethyl Sulfoxide, Sigma-Aldrich Canada Co., Oakville, Ontario, Canada). A total volume of 0.25 µl of quinolinic acid was infused at the rate of 0.175 µl per min in each site using a 30-gauge injection cannula attached to a 10 µl Hamilton syringe via polyethylene tubing (PE-50). The injection needle was left in place for 2 min following the injection to allow diffusion of the drug. The scalp incision was then closed with sutures. Rats were given subcutaneous injections (0.02 mg/kg) of meloxicam (Boehringer Ingelheim, Germany) and monitored for 24 hr before returning them to the vivarium. The animals recovered in their home cages (pair housed) for one week before resuming behavioural testing.

At the end of behavioural testing, all subjects received lethal injections of sodium pentobarbital (100 mg/kg i.p.) and were perfused with physiological saline and 4% paraformaldehyde. The brains were post-fixed for 24 h in PFA and then transferred and stored in 30% sucrose and PBS with sodium azide (0.02%) for a minimum of 48 h before sectioning. The brains were sectioned in the coronal plane at 40 µm thickness using an SM2010R freezing microtome (−19°C, Leica, Germany). Every second section through the region of interest was wet-mounted on glass microscope slides and later stained with cresyl violet. Images of sections were digitized using a NanoZoomer (Hamamatsu, Japan) and evaluated for lesion quality.

Behavioural analysis

We quantified several behavioural measures in the CCT. EFS was defined as the trials where the animals sampled both feeders after making an entry into the poke port (Figure 6A EFS panel). The probability of lose-shift was calculated as the probability that the rat would shift feeder choice in the consecutive trial following reward omission. Likewise, the probability of win-stay was calculated as the probability that the rat would repeat the selection of the same feeder on trials immediately following rewarded trials. The number of trials represents the total number of complete trials within a session. Only sessions with more than 100 trials were included in the analysis, which impacted only the analysis of behaviour in the rats with lesions to the DLS (1 session out of 37 was excluded). The calculation of the percent of rewarded trials (wins) represents the percentage of all complete trials in which the rat was reinforced with sucrose. Response time measures the time taken to reach the feeder after the exit of poke port, whereas inter-trial interval (ITI) was defined as the time between the first exit of the reward feeder and the next entry into the poke port. Infrared beam break detectors in the feeders were used to detect the number of anticipatory licks during the short hardware-determined delay (typically 200 - 600 ms) before reward delivery (licks).

Data were analyzed with MATLAB (version R2013a; MathWorks, MA, USA) and SPSS (version 21.0; IBM, NY, USA). Analysis of variance (ANOVA), Repeated-measures analysis of variance (RM-ANOVA), and mixed ANOVA was used to assess the significance of lesion on behavioural measures ($p < 0.05$). Where the main effects were statistically significant, a post hoc Tukey or Bonferroni test was used to determine which marginal means differed significantly.

Results

Rats were required to perform a very brief (100 ms) nose poke and then locomote to one of the two adjacent reward feeders for the possibility of receiving sucrose solution as a reward (Figure 6*A)*. The optimal behavioural sequence for maximizing the number of rewards on the task is to commit a nose-poke in a centrally located port, enter one randomly-chosen feeder, and then begin the next trial by committing a nose-poke in the port. Locomoting to the alternate feeder (i.e. EFS) without committing the nose poke is never reinforced and has both effort and opportunity costs. We initially suspected that animals would be more likely to approach the alternate feeder following reward omission, as compared to reward delivery. However, we found no significant difference in the probability of EFS after a win versus after a loss in well-trained animals in cohort 1 (paired t-test; $t_{67} = 0.96$, $p = 0.34$; Figure 6*B*).

We next sought to discern if EFS affected animals' choices on subsequent trials. A computer chose the well to be baited on each trial according to each rat's past actions and reinforcements such that the optimal choice strategy by the rat is a random selection. Nonetheless, most rats tend to engage in the non-optimal strategy of lose-shift responding above chance levels (i.e. more than 50% of trials). Our previous work showed that there are several variables that can affect choice on this task. Importantly, the probability of lose-shift responding strongly decays with increasing inter-trial interval (ITI) between the time of reward omission and the start of the next trial on this task (Chapter two; Gruber and Thapa 2016). This relationship is also present in the current data (black dots in Figure 6*C*). The EFS behaviour increases the ITI because of the additional time it takes to locomote to the alternate feeder prior to the subsequent nose-poke. The ITI distributions for trials following EFS (EFS+) is therefore shifted from that of trials not following EFS (EFS-). We, therefore, limited the subsequent analysis of lose-shift responding in this cohort to trials with ITI in the range of 3-8 s to ensure sampling from both EFS+ and EFS- trial types throughout the ITI range. The probability of lose-shift is strongly decreased following trials with EFS for all ITI in the test range (green circles in Figure 6*C*). We hypothesized that this could result from the animals using a lose-shift response from the last feeder sampled in the trial (rather than the first to be sampled). This is strongly supported by two analyses. First, the mean probability of lose-shift for each rat is significantly higher when computed after removing trials following EFS (i.e. mean for the EFS+ type) than for the mean computed with all (EFS+ and EFS-) trials (t_{67} = 9.1, *p* = 1.00E-6 or less; Figure 6*D*). If the EFS had no effect on subsequent choice, then removing these trials should have had no effect on the mean. Second, the mean lose-shift responding for each rat computed over all trials (EFS+ and EFS-) based on the last feeder visited is much higher than the mean computed from the first feeder visited (t_{67} =10.1, *p* = 1.00E-6 or less; Figure 6*E*). In other words, animals based their lose-shift strategy on the last feeder visited, regardless if this was during a trial or not. This suggests the neural systems involved in this decision-making process mistakenly expected a reward at the second feeder and is consistent with the characterization of lose-shift responding as a ‘choice-reflex’ (Chapter two; Gruber and Thapa 2016). The large effect of EFS on choice motivated us to further investigate its properties and neural basis.

Rats engaged in EFS on nearly 50% of trials in the first few sessions, but this significantly decreased with training (RM-ANOVA, main effects of the session: $F_{7,30}$ = 48.95, p = 1.00E-6; Figure 7*A*). However, the EFS responses persisted at substantial levels (mean = 0.230 +/- 0.106) even after extended training (8 sessions after training was complete). We next sought correlational evidence whether the neural systems promoting EFS are associated with those promoting either win-stay or lose-shift responding, which have distinct properties and neural dependencies (Chapter 2; Skelin et al. 2014; Gruber and Thapa 2016). We excluded all trials following EFS in the subsequent analysis of win-stay and lose-shift responding to avoid the immediate effect of EFS on choice. We examined the session-averaged responses of each rat on the last day of testing (8th session). The rats showed a probability of lose-shift (mean = 0.692 +/- 0.020) that was higher than chance levels (p = 0.50), consistent with previous reports (Chapter two; Gruber and Thapa 2016). Lose-shift did not decrease over the training/testing sessions (RM-ANOVA: $F_{1,\ 36}$ = 0.531, p = 0.471; Figure 7*B*). Conversely, the animals showed a lower than chance probability of win-stay on the last day of testing (mean = 0.395 +/- 0.013), and this again is stable across the training/testing sessions (RM-ANOVA: $F_{7,30}$ = 0.427, p = 0.877; Figure 7*C*). We next tested for relationships among these behavioural measures. EFS showed no significant linear correlation with win-stay ($F_{1,\ 67}$ = 1.5, p = 0.220; r^2 = 0.02; Figure 7*D*) or lose-shift ($F_{1,\ 67}$ = 3.5, p = 0.067; r^2 = 0.05; Figure 7*E*) responding, but win-stay was negatively correlated with lose-shift responding ($F_{1,\ 67}$ = 34.4, p = 1.00E-6; r^2 = 0.34; Figure 7*F*). This suggests that win-stay and lose-shift are opponent processes and/or have distinct temporal sensitivities, whereas EFS prevalence is independent of both under normal conditions.

We next wanted to assess if EFS or the other response variables varied within sessions. EFS responses significantly decreased during the session (F4,64 = 37.46, p = 1.00E-6 or less; Figure 7G). In contrast, neither lose-shift nor win-stay responding varied within session (lose-shift: F1, 4 = 7.3, p = 0.07; win-stay: F1, 4 = 1.9, p = 0.26; Figure 7H). The dissociation of these within-session variances further indicates that EFS is distinct from the neural mechanisms of lose-shift or win-stay responding. The reduction of EFS during the session could be due to changes in either motivation (e.g. thirst) or task uncertainty, which are both expected to decrease as the session progresses. These, however, should diverge with training such that uncertainty should decrease as experience accumulates across sessions, whereas motivation for reward should be relatively invariant among sessions. We, therefore, examined how EFS decreased within the session as a function of experience (training sessions) in a new group of male LE rats with extended training (Cohort 2; n = 30). There was a main effect of the training session (F3,84 = 45.6, p = 1.00E-6 or less) and of trial in the session (F9,252 = 27.635, p = 1.00E-6 or less), as well as a significant trial*session interaction (F27,756 = 3.34, p = 0.001). The within-session decrease became smaller with increased training (Figure 7I) but was still significant at the 18th session (F9,261 = 4.018, p = 1.00E-6 or less). These correlational data support the hypothesis that it is task familiarity rather than motivation that drives EFS. We next sought direct evidence for this hypothesis.

In order to discern if the EFS is promoted by the motivation for the reward, as would be expected by phenomena driven by Pavlovian systems, we conducted a devaluation experiment in Cohort 2 after 12 sessions of training. Animals were allowed to drink a fixed amount of liquid sucrose prior to the task, in a counterbalanced design. This should decrease EFS if it is promoted by the motivation for the outcome. Pre-feeding decreased the number of trials completed in a volume-dependent manner (RM-ANOVA, main effect: $F_{2,46} = 35$, p = 1.00E-10; Figure 8A) but had no effect on the number of trials with EFS ($F_{2,46} = 2.4$, p = 0.10; Figure 8B). Thus, the relative rate of EFS to operant responses increased with devaluation (RM-ANOVA with Greenhouse-Geisser correction: $F_{1.9,43} = 6.7$, p = 0.003; Figure 8C). This was unexpected, and we wanted to test whether this could be an artifact of an unplanned factor within our control. We, therefore, replicated the experiment under conditions of the increased variance of originally unplanned factors. The replication was conducted by new investigators (female instead of male), at a different time of year, and with a new heterogeneous group of rats (Cohort 3; n = 52) that included male LE (n = 16), female LE (n = 6), transgenic female LE (n = 15) and transgenic male LE (n = 14) with an inert transgene (see METHODS). This cohort was bred in our facility, whereas Cohort 2 was shipped from a commercial breeder. Despite these changes, the results were remarkably similar to the first devaluation experiment. Devaluation again decreased trial completion ($F_{1.6,43.8} = 51.0$, p = 1.00E-6; Figure 8D) but not EFS ($F_{2,50} = 1.0$, p = 0.36; Figure 8E), yielding an increased relative rate of EFS ($F_{1.64,41} = 8.0$, p = 0.002; Figure 8F). Note that the rate of EFS is higher in this group (Cohort 3) as compared to that in Cohort 2 because they had fewer training sessions prior to the devaluation. These data provide strong evidence that EFS is a robust phenomenon independent of outcome valuation.

We next tested if uncertainty would affect the relative EFS rate. We allowed rats (n = 16 male LE wild-type from Cohort 3) to perform the task for 100 trials with their customary 13 cm barrier separating the nose-poke from the feeders. We then took the rats out of the box and replaced the barrier with a longer one, a shorter one, or one the same length. Rats were then placed back in the box and allowed to perform an additional 100 trials. The relative EFS rate increased for either novel barrier length as compared to the familiar one (RM-ANOVA, time*barrier: $F_{14,294} = 3.34$, $p = 1.00E-5$; Figure 9). These data indicate that EFS is not related to the effort of circumnavigating the barriers because we would then expect a monotonic length-EFS relationship rather than a parabolic one. These results indicate that a change in the apparatus is sufficient to transiently increase EFS, suggesting that EFS is promoted by uncertainty about the task or apparatus.

The previous data indicate that EFS is not sensitive to outcome devaluation and therefore not likely directly affected by Pavlovian associations. EFS could instead arise from the inability to suppress motor responses leading to the feeders. Such impulsive actions are typically associated with processing in the sensorimotor regions of the rodent caudate-putamen in the dorsolateral striatum (A. M. Graybiel, 1998), which do not show devaluation effects (Balleine, Delgado, & Hikosaka, 2007). If so, then damage to this region would be expected to reduce the rate of EFS. We tested this by producing bilateral excitotoxic lesions of either the dorsolateral striatum (DLS; n = 7) or the nucleus accumbens core (NACc; n =

7) and comparing the resultant CCT behaviour to control animals (n = 7) from the same cohort. The location and extent of the lesions (Figure 10A-B) are similar to previous reports from our group and others (Hall, Parkinson, Connor, Dickinson, & Everitt, 2001; Skelin et al., 2014).

The DLS-lesioned rats had higher response times than controls ($F_{2,16}$ = 19.4, p = 1.00E-6; Figure 10*C*) but had equivalent percentages of rewarded trials compared to controls ($F_{2,16}$ = 1.0, p = 0.4; Figure 10*D*). They showed above normal amount of licking in the feeder, suggesting no motivational deficit (Figure 5*E*). The DLS rats had a much lower rate of trial completion than controls (ANOVA Main effect: $F_{2,15}$ = 16.4, p = 2.00E-4; Tukey post-hoc shown in Figure 10*F*). Their rate of EFS was not statistically different than controls but tended to be higher ($F_{2,15}$ = 2.8, p = 0.09; Figure 10*G*). The relative rate of EFS to operant responses was therefore significantly higher in DLS-lesioned animals than controls ($F_{2,15}$ = 22.9, p = 5.00E-5; Figure 10*H*). The NACc-lesioned rats were not different from controls in either trial completion or EFS (post-hoc shown in Figure 10*F-H*). These data indicate that EFS does not depend critically on either striatal region, and further suggests that EFS is not a product of impulsive engagement of habits dependant on the DLS.

Further evidence that EFS is independent of these striatal regions comes from the dissociation of lesion effects on EFS from win-stay or lose-shift responding. Consistent with our previous finding (Skelin et al. 2014), the DLS lesion group made significantly fewer lose-shift responses than the control or NACc-lesion groups ($F_{2,16}$ = 15.83, p = 1.00E-6; Figure 11*A*), and this reduction was irrespective of the ITI (Figure 11*B*). The DLS-lesioned group had a lose-shift response probability at chance levels for all ITI values. The NACc lesion group, in contrast, showed a higher probability of lose-shift than controls across the range of the ITI (Figure 11*B*). Furthermore, the large reduction in lose shift in DLS-lesioned animals (compared to controls) is also evident when including EFS+ trials and computing lose-shift from the last feeder sampled (controls = 0.65 ± 0.001; DLS-lesioned = 0.40 ± 0.01; $t_{10} = 4.0$, $p = 0.003$). The effects of lesion location on win-stay responding had an inverse relationship; the NACc-lesioned group showed a marginally significant reduction in win-stay compared to the other groups (ANOVA main effect: $F_{2,16} = 3.782$, $p = 0.045$; Figure 11*C*), whereas DLS lesions had no reduction in win-stay (post hoc Tukey, $p = 0.996$). This reduction occurred over the range of the ITI (Figure 11*D*), suggesting that it normally plays a role in suppressing such actions, whereas the NACc lesion group showed a non-significant trend for increased EFS compared to controls (post hoc Tukey, $p = 0.061$). In sum, lose-shift responding depends on the integrity of the DLS, whereas win-stay depends on the NACc. The number of EFS events was not reduced by either lesion and in fact showed a non-significant trend to increase in lesioned animals, whereas the ratio of EFS to operant task performance was much higher in DLS-lesioned animals than controls.

Discussion

Decision-making is a complex process influenced not only by the drive to maximize cumulative reward but also by proximate influences such as the drive to approach feeders, outcome-related cues, and ‘choice reflex’ tendencies like lose-shift and win-stay responses. These influences likely involve interactions among multiple brain circuits with unique information processing capacities (Daw et al. 2005; Balleine and O'Doherty 2010). Here, we have revealed dissociations among regions of the striatum in win-stay, lose-shift, and the suppression of approach to the feeders outside of the normal task sequence (e.g. context). This latter behaviour (EFS) was insensitive to reinforcements, but it strongly affected subsequent choice in the task; rats lose-shifted away from the last feeder sampled prior to the subsequent nose-poke, regardless if feeder entry was from a choice within the operant task or if it was a consequence of EFS. This is a novel mechanism by which reinforcement-driven task performance could be modulated indirectly by manipulations that affect approach behaviours outside of the task context.

The EFS behaviour never fully diminished despite the lack of any positive reinforcement (Figure 7*I*). EFS occurred in control animals about a quarter of their trials even after extended training. A similar phenomenon was observed by R. A. Boakes (1977) in his study of goal tracking and sign tracking behaviours when he introduced reward omission conditions. His omission contingencies were effective in reducing the frequency of the goal-tracking response, although it rarely eliminated them. Boakes interpreted the failure to diminish responses with reward omission as an indication that the goal-tracking and sign-tracking responses are in competition for behavioural control. We speculate that similar opponent influences result in the persistence of EFS in the CCT. One of these processes drives the instrumental responding and involves the DLS, as evidenced by the reduction in trial completion after lesion of this structure. We have no evidence to suggest what process promotes EFS in the present task.

Although there are no explicit discriminative stimuli predicting reward delivery in our task, we cannot rule out the formation of associative learning involving implicit stimuli. These could involve stimulus-outcome (S-O) or response-outcome (R-O) contingencies when the rat is reinforced at the feeder. Indeed, the use of multiple outcomes and lack of discriminative stimuli promote R-O and/or S-O control (Holland 2004). It is possible that rats break the operant response into multiple components. If one of these represents entry of the lane to the feeder, it is possible that the R-O of this portion gains strength during training. However, this suggests the EFS should increase with training, whereas the data reveal that it decreases. Alternately, the feeder could have gained incentive salience because it is the most proximal conditioned stimuli (CS) to the unconditioned stimuli (UCS, i.e., sucrose). Rats, therefore, may be motivated to make an EFS response due to Pavlovian (S-O) attraction to stimuli proximal to the UCS. The main problem with such an interpretation is the fact that the absolute rate of EFS trials was not reduced by the devaluation of the outcome via pre-feeding in either of two distinct cohorts. These data suggest that EFS is driven by associations other than R-O or S-O. An alternative mechanism could be stimulus-response (S-R) responding, which is largely unaffected by devaluation and is thought to involve DLS (Dolan & Dayan, 2013; Graybiel, 1998; Yin & Knowlton, 2004; Yin et al., 2004). However, the rate of EFS was not reduced by lesions of the DLS in the present study, suggesting the involvement of some other brain region. An obvious candidate is

NACc. Dopamine depletion in this structure drastically reduce engagement in instrumental responding (Nicola 2010), and NACc neurons encode nearby manipulanda and presumably support approach (Morrison et al. 2015). Moreover, we previously found that infusion of amphetamine into NACc increased EFS (Wong et al. 2016), consistent with reports that this increases Pavlovian conditioned approach (Parkinson et al. 1999; du Hoffmann and Nicola 2014). It was thus surprising that lesions of NACc in this study did not decrease EFS. Perhaps the extent of lesions was insufficient, or some other brain region can quickly take over the NACc's contribution to EFS. Nonetheless, this is consistent with proposals that multiple reinforcement learning and memory systems can compete for control of behaviour (Dayan et al. 2006).

Is the shuttling between feeders (EFS) simply an error reflecting incomplete mastery of the task contingencies, or does it reveal something about ingrained foraging behaviours in rats? We argue that it is the latter. EFS does not fully extinguish after extensive training and appears to increase at times of less certainty of the task: initial training; the beginning of sessions; and following a switch of the barriers. Its insensitivity to both devaluation and to reward outcome (wins/losses) indicates that EFS is not driven by motivation, frustration, or outcome expectation. We, therefore, speculate that EFS may serve a role in ethological contexts to increase explorative actions. Reinforcement theory indicates this is a good policy in environments with uncertainty (Sutton and Barto 1998; Kakade and Dayan 2002; Sugrue et al. 2004; Daw et al. 2006). We argue that the natural environment involves sufficient variability in such a large state space that animals will always face some level of uncertainty about features pertinent to survival. We speculate that the rodent brain may, therefore, have evolved a system that promotes exploration for foraging, particularly at times of uncertainty or when opportunity costs are low. Moreover, the neural systems promoting exploration may be inhibited as those that promote exploitative actions gain associative strength. This would account for the reduction of EFS with training, and its tendency to increase following striatal lesions in well-trained animals.

A striking and unexpected feature of the data is that the feeder approach during the inter-trial-interval strongly affected subsequent choices on task. We observed that EFS triggered the lose-shift response, suggesting that the reward error signal conveyed to this system treats EFS similar to the operant approach during the task. This lack of context may be explained by the properties of the DLS. We have shown previously (Skelin et al. 2014) and here (Figure 11*A-B*) that lose-shift depends on the DLS, and this structure is generally not contextually sensitive (McDonald and White 1993). The ability of EFS to trigger lose-shift responding reveals cross-talk between behavioural control systems that, to our knowledge, has not been previously described. This could be related to proposals that reward prediction error signals in the striatum are ‘factored’ in order to account for complexity in the world, and go on to impact multiple reinforcement learning systems (Lesaint et al. 2014). The effect of EFS on subsequent choice shown here highlights the need to consider actions prior to trial initialization when analyzing the effects of treatments on decision-making.

EFS is modulated by drugs such as D-amphetamine (Wong et al. 2016), but not others such as Δ-9-tetrahydrocannabinol (Wong et al. 2017). Moreover, it appears to be sexually dimorphic in rats and may be subject to modulation by stress, inflammation, or other factors (unpublished observations). Such effects on EFS, and the effect of EFS on subsequent choice highlight the need to consider actions before trial initialization when analyzing the effects of treatments on decision-making.

Chapter four

Lesions of Lateral Habenula Attenuate Win-stay but not Lose-shift Responding in an Operant Binary Choice Task[3]

Abstract

Multiple neural systems contribute to choice adaptation following reinforcement. Recent evidence suggests that the lateral habenula (LHb) plays a key role in such adaptations, particularly when reinforcements are worse than expected. Here, we investigated the effects of bilateral LHb lesions on responding in a binary choice task with no discriminatory cues. LHb lesions in rats decreased win-stay responses but surprisingly left lose-shift responses intact. This same dissociated effect was also observed after systemic administration of d-amphetamine in a separate cohort of animals. These results suggest that at least some behavioural responses triggered by reward omission do not depend on intact LHb or dopamine signalling.

Introduction

Animals often repeat choices that lead to reward and forgo behaviours that are unproductive or result in negative outcomes (Thorndike 1927). This choice behaviour may be influenced by multiple response strategies, such as reinforcement learning through the integration of recent reinforcement history over several trials, or a heuristic such as lose-shift or win-stay that disproportionately weighs immediately previous outcomes (Chapter two; Gruber and Thapa 2016).

[3] Submitted to Neuroscience Letters as: Thapa, R., Wong, S.A., Sutherland, R.J., and Gruber, A.G. (2017). Lesions of lateral habenula attenuate win-stay but not lose-shift responding in an operant binary choice task. Student's Contribution: performing surgeries, data collection, analysis, and writing the manuscript.

Such adaptive behavioural strategies are thought to be mediated by dissociated circuits involving multiple brain structures (Schultz et al. 1997; Gruber and McDonald 2012; Cohen et al. 2015). Recent findings indicate that a small nucleus in the epithalamus, the lateral habenula (LHb), plays a crucial role in a reinforcement-driven choice adaptation by encoding choice outcomes on a trial-by-trial basis (Matsumoto and Hikosaka 2007; Baker et al. 2015; Kawai et al. 2015; Mizumori and Baker 2017). Negative outcomes such as aversive stimuli or reward omission (Salas et al. 2010; Kawai et al. 2015; Tian and Uchida 2015) activate LHb neurons, which in turn inhibit most of the midbrain dopamine neurons in the substantia nigra (SN) and the ventral tegmental area (VTA) (Christoph et al. 1986; Ji and Shepard 2007). This effect is mediated by GABAA receptors on the dopamine neurons (Ji and Shepard 2007) through both direct excitatory LHb projections onto GABAergic interneurons near the dopamine neurons (Brinschwitz et al. 2010), as well as indirect projections via the rostromedial tegmental nucleus (Balcita-Pedicino et al. 2011; Goncalves et al. 2012; Brown et al. 2017). The habenular inhibition of dopamine neurons causes phasic suppressions of dopamine levels in downstream structures such as the striatum (Sugam et al. 2012) that is thought to encode a negative reward prediction error(Montague et al. 1996). The striatal efferent neurons express dopamine receptors, providing a mechanism by which the LHb can influence adaptive choices relative to recent outcomes (Bromberg-Martin et al. 2010; Schultz 2013; Baker et al. 2016).

Behavioural results from several previous studies indicate that intact LHb is required for flexible responding on a trial-by-trial basis in the face of changing associations between outcomes and cues or contexts (Thornton and Davies 1991; Baker et al. 2015; Kawai et al. 2015). For instance, rats with inactivated LHb have difficulty in flexibly biasing responding to locations with a higher reward probability when the probabilities of reward are reversed among two locations (Baker et al. 2015). It remains unclear, however, whether this reversal learning deficit is due to an inability to learn from negative outcomes, positive outcomes, or both. Furthermore, most behavioural studies of LHb have focused on tasks with cues informative of outcome contingencies that are unequal among the choice options. It is unknown whether the same LHb circuitry is involved in rapid win-stay or lose-shift responses when there are no predictive outcome cues and when the reinforcements are unpredictable.

In the present study, we investigated the contribution of the LHb to behavioural responses in an uncued binary choice task in which the optimal response strategy is a random choice. This allowed us to not only investigate decision-making in an unpredictable environment but also to parse the immediate influence of wins and losses on decisions in the absence of a strong feeder preference. Because LHb lesions are expected to increase dopamine levels and attenuate the negative reward prediction error (Bromberg-Martin et al. 2010), we compared the behaviour of lesioned animals to a separate cohort that received pre-testing injections of amphetamine, which also increases dopamine levels. We hypothesized that disruption of the ability to generate reward prediction error signals through phasic reductions in dopamine levels caused either by the lesions of LHb or administration of amphetamine, would attenuate both lose-shift and win-stay responses.

Methods

Subjects

Forty male Long-Evans rats and another separate cohort of eight males, (Long-Evans, Charles River, Quebec, Canada) weighing between 450 and 600 g were used in experiments 1 and 2, respectively. All animals were housed in pairs in a transparent plastic cage with corn-cob bedding. The experimenter handled the rats daily for one week before the beginning of training. During behavioural training and testing, access to water was limited to one hour per day but was unrestricted otherwise. The vivarium was maintained at 21°C and 12-h light/dark cycle (lights off at 7:30 pm). All experimental procedures were approved by the University of Lethbridge Animal Welfare Committee and adhered to the guidelines of the Canadian Council on Animal Care.

Competitive Choice Task (CCT)

Behavioural testing took place in custom-built aluminum boxes (26 X 26 cm). Each box contained two cue lights and two liquid delivery feeder wells on either side of a central nose-poke port (Figure 12A). Infrared emitters and sensors in the feeder wells and central port detected animal entry. Following the illumination of the cue lights, the rats poked their snout into the central port to initiate a trial and then responded by going to one of the two feeder wells. A 13-cm-long aluminum barrier orthogonal to the wall separated each well from the central port. This added a choice cost and reduced choice bias originating from body orientation. The control of the behavioural task was automated with a microcontroller (Arduino Mega) receiving commands via serial communication from custom software written in MATLAB. We attempted to reduce acoustic startle from sounds outside of the testing chamber by playing a local radio station as a constant background audio stimulus.

Trials of the task began with the illumination of two cue lights. Animals then had 15 seconds to commit a nose-poke into the central port, and subsequently respond to one of the two possible feeder wells (Figure 12C). The computer selected a priori which feeder well to reward based on a ‘competitive’ algorithm that used the well choice and reinforcement history of the rat to predict the choice in the current trial and thereby minimize the number of rewarded trials (Lee et al. 2004; Skelin et al. 2014). If the rat selected this rewarded feeder, it received a 60 µL drop of 10% sucrose solution (win). If the rat chose the non-rewarded well, the feeder was left empty (lose) and the house light was illuminated. The competitive algorithm, therefore, punished predictable response patterns. The optimal solution for the rat was to be as random as possible in its feeder well choices. Daily sessions of the task were 45 minutes in duration.

All animals were trained on the competitive choice task by gradually introducing components of the task. Initially, there were no barriers between the central port and feeder wells, and every response was rewarded. Any obvious feeder preference was corrected by briefly (10-20 trials) introducing a feeder cover to prevent access to the preferred side. Once rats performed 100 trials (typically in the first session), the session was terminated, and subsequent sessions used the competitive algorithm. The barrier separating the nose-poke port and feeder wells was increased in discrete lengths (4, 8, and 13 cm) over several sessions (typically 4 - 5). The shaping was complete when the animal's performed over 100 trials with the 13-cm barrier within the 45 minutes session over two consecutive days.

Experiment 1: LHb lesion surgery

Once the animals were trained in CCT, they were randomly assigned to either the LHb lesion (n = 27) or control group (n = 13). Before the start of the surgery, all rats received Buprenorphine (Alstoe Ltd., UK) for analgesia. The animals were anesthetized using 4% isoflurane gas (Benson Medical Industries Inc., Ontario, Canada) in oxygen flowing at 1.0 L/min. The surgical plane was maintained with 2% isoflurane throughout the surgery. The animals were mounted on a stereotaxic frame (Kopf Instruments, Tujunga, CA, USA) and a midline incision was made to expose the skull. Burr holes were drilled through the skull and the lesion electrodes were lowered to the following coordinates from bregma [in mm (AP, ML, DV)]: -3.8, 0.6, -4.4]. Lesion electrodes were custom built with 0.26 mm diameter dummy guide cannula insulated with liquid electrical tape (Permatex, item #85121, Hartford, CT, USA) except 0.5 mm exposed wire at the tip. Lesions were made with a stimulus isolator (World Precision Instruments, Sarasota, FL, USA) by passing anodal direct current of 1.2 mA for 15-sec between the electrode placed stereotaxically in the selected area and another electrode attached by alligator clip to the tail wrapped in saline-soaked cotton gauge. Control animals received the same surgical procedure except no current was passed through the lesion electrode.

Experiment 2: Systemic amphetamine injections

A separate cohort of animals (n = 8) extensively trained in CCT was used in the second experiment. D-amphetamine hemisulfate (Sigma—Aldrich, Oakville, Ontario, CAN) was dissolved in 0.9% saline in 1.5 mg/kg concentration. The animals received intraperitoneal (IP) injection of either saline or 1.5 mg/kg dosage of amphetamine in a counterbalanced repeated measures design 15 minutes before testing. This occurred over three days with a non-injection day in between to allow for drug wash-out. Prior to the counterbalanced design, animals underwent a day of saline injections followed by the task to habituate them to the process. This was not used for analysis. Only two dosages were used in this study based on the effective dosage seen in our prior study (Wong et al. 2016).

Histology

At the end of the experiment, all subjects received lethal injections of sodium pentobarbital (100 mg/kg i.p.) and animals from the LHb lesion experiment were perfused with physiological saline and 4% paraformaldehyde. The brains were post-fixed for 24 hr in PFA and then transferred and stored for another 48 hr in 30% sucrose and PBS with sodium azide (0.02%). The brains were sectioned using a freezing microtome (−19°C, Leica, Germany) in the coronal plane at 40 μm thickness. Every second section through the region of interest was wet-mounted on slides and later stained with cresyl violet.

LHb damage quantification

Given the potentially contrasting roles of medial and lateral habenula to choice (Viswanath et al. 2013), we used a Cavalieri method to compute stereological volume estimates of the percentage of LHb damage (Schmitz and

Hof 2005) and selected only the animals with significant lesions limited to the LHb. Images of cresyl violet stained sections were taken using a Zeiss epifluorescent scope (ZEISS, Germany) attached to a CCD camera. Images were then analyzed using ImageJ software. A sampling grid size of 80 µm was created and randomly thrown over each image and the total number of points in contact with the LHb tissue in each section was counted. The software estimated the total volume of the spared LHb by summating the per section product of the number of points in contact with tissue, the section thickness, and the section-sampling fraction. Percent LHb damage in each of the damaged rats was calculated by dividing the quantified spared tissue volume by the average LHb volume of five animals from the control group and then multiplying by 100.

Statistical analysis

Data were analyzed with MATLAB (version R2013a; MathWorks, MA, USA) and SPSS (version 21.0; IBM, NY, USA). One-way analysis of variance (ANOVA), repeated measures ANOVA, and mixed model ANOVA including a within-subjects factor was used to assess the significance of lesion and systemic amphetamine injections on the behavioural measures ($p < 0.05$).

Only sessions with more than 100 trials were included in the analysis. The probability of lose-shift was calculated as the probability that the rat would shift feeder choice in the consecutive trial following reward omission (Figure 12C). Likewise, the probability of win-stay was calculated as the probability that the rat would repeat the selection of the same feeder on trials immediately following rewarded trials. The number of trials represents the total number of complete operant trials within a session. Extraneous feeder sampling (EFS) was defined as the trials where the animals sampled both feeders after making an entry into the poke port. The two feeder-well locations and the central port were defined as three distinct foraging patches and the measure patch visits were computed as the total number of distinct visits of any one of the location based on infrared emitters and sensors that detected entry—this indirectly measures the amount of locomotion within a session. Response time was defined as the time taken for the rat to locomote from the nose-poke port to one of the feeders. Infrared beam break detectors in the feeder wells were used to detect the number of licks. The measure total licks summate the number of beam breaks during a specified time interval. The calculation of the percent of rewarded trials (wins) represents the percentage of all complete trials in which the rat was reinforced with sucrose.

The LHb and its effect on striatal neurons through its inhibitory control of dopaminergic neurons provide a potential circuitry for producing adaptive responses to reinforcement outcomes (Bromberg-Martin et al. 2010; Schultz 2013; Baker et al. 2016). We can investigate the viability of this hypothesis by observing the behavioural effects of manipulations of parts of this circuitry in the same behavioural task. The present paper reports result of LHb lesions and the effects of systemically elevating dopamine levels on prediction error encoding by the dopamine neurons. We have previously reported the behavioural results of amphetamine microinfusions into the nucleus accumbens core of the striatum (Wong et al. 2016). Long Evans rats were used in all three studies, and the same training and testing protocol of the competitive choice task was used. We, therefore, predicted similar behavioural effects across the three studies. To compare the behavioural effects between the studies we computed Δ win-stay, Δ lose-shift, and Δ response-time by subtracting the means for each data set from their own respective controls and plotting the results on the same graph for each variable (Figure. 16).

Results

Histological verification and stereological quantification of LHb lesions

We estimated the total volume of an intact LHb to be on average 0.325 mm^3 based on stereological volume estimates in five of the control rats. We made bilateral focal lesions of the LHb by passing electrical current. Any rats with incomplete LHb damage (less than 45%) in either hemisphere or lesions that extended to medial habenula or too deep into the thalamus, were excluded. This selection was conducted prior to group-wise analysis of data. Only 8 of the 27 rats receiving LHb lesions were included in the analysis. This large attrition rate owes to the difficulty of generating large, selective bilateral lesions of the LHb. The LHb-lesioned rats included in the analysis (n = 8) had an average of 58.0% LHb damage (SD = 14.9; Min = 45.1; Max = 86.1). Figure 12B shows representative LHb damage.

Experiment 1: Behavioural results of LHb lesions

Rats were trained to make repeated binary choices in the competitive choice task. There were no explicit outcome cues informing the animals which feeder well was baited. Rather, a competitive algorithm kept track of the animal’s previous reinforcement history and baited the least likely feeder in each trial. This discouraged the formation of any side bias or patterns of responding. In other words, to collect maximum reward the animals were forced to distribute their choices as randomly as possible. Rats, however, generate lose-shift and win-stay responses on this task at a rate greater than chance, even-though these are non-optimal (Chapter two; Skelin et al. 2014; Gruber and Thapa 2016). We predicted that LHb lesions would disrupt the reward prediction error signals and attenuate the animals' sensitivity to reinforcement, thereby reducing lose-shift responding and possibly win-stay responding.

As expected, the control animals' choices were largely influenced by wins and losses in the preceding trial (Figure 13). The animals with LHb lesion were no different than controls in the tendency to make lose-shift choices (F(1,19) = 0.062, p = 0.806; Figure 13A), and the probability of lose-shift showed no correlation with the percentage of LHb damage (r^2 = 0.056, p = 0.811; Figure 13B). Moreover, the dependence of lose-shift on the ITI was not different between groups (stats, Figure 13C). In contrast, the probability of win-stay responses was reduced significantly in the lesioned animals as compared to controls (F(1,19) = 9.542, p = 0.006; Figure 13D), and a correlation analysis between lesion size and the probability of win-stay showed that the larger the LHb lesion, the greater was the attenuation of win-stay prevalence (r^2 = -0.582, p = 0.006; Figure 13E). This difference appears to hold over the relevant range of ITI (Figure 13F). In sum, these results indicate that LHb lesions only attenuate immediate responses following reward delivery (wins) and not to reward omissions (loss) in this task.

LHb lesions are known to produce hyperactivity (Murphy et al. 1996), and we observed this effect on locomotion in our LHb lesioned cohort. The LHb lesioned animals completed significantly more trials than the controls during sessions of the same duration (F(3,172) = 4.94, p = 0.002; Figure 14A). Despite extended training in the competitive choice task, the rats would sometimes shuttle directly between the feeder wells and make orofacial contact with them without first initiating a trial with a nose-poke in the central port; we termed this behaviour extraneous feeder sampling (EFS). The lesioned animals showed a statistically greater tendency to make EFS responses than the controls (F(1,19) = 18.9, p < 0.000; Figure 14B). A trial with EFS was not counted towards the total number of trials completed, but it did indicate locomotion of the rats. To quantify the overall extent of locomotion in the operant box, we divided the box into three foraging patches—the centre nose poke port and the two feeders. The count of the discreet visits to each section, or 'patch visits', gave an overall measure of locomotion. The LHb-lesioned animals showed a significant increase in the number of patch visits compared to the controls across the session (F(4,186) = 2.94, p = 0.026; Figure 14C). In addition, the lesioned animals were also much faster than controls in their response time (F(1,47) = 13.1, p = 0.001; Figure 14D). Nonetheless, the two experimental groups were similar in the total number of feeder well licks (F(4,206) = 0.492, p = 0.758; Figure 14E) and the total percentage of wins (F(5,281)= 0.685, p = 0.646; Figure 13F). These results indicate that the LHb lesions in the experimental group induced motoric hyperactivity but not motivational deficiencies.

Experiment 2: systemic amphetamine injections

We used a repeated measures design to test the effects of two dosages (0 and 1.5 mg/kg) of acute d-amphetamine (AMPH) administration on choice behaviour in the task in a cohort of extensively trained animals (Figure 15). Lose-shift responding did not change with AMPH administration (F(1,4) = 0.617, p = 0.476; Figure 15A). However, win-stay responding decreased significantly (F(1,4) = 21.92, p = 0.009; Figure 15B). This pattern is analogous to that of LHb lesions. The animals had a significantly faster response time when on amphetamine (F(1,4) = 9.763, p= 0.035; Figure 15C). However, the percentage of wins was not affected (F(1,4) = 0.908, p = 0.395; Figure 15D). The animals on AMPH showed a greater tendency for motoric hyperactivity as measured by number of trials (F(1,4) = 5.293, p = 0.083; Figure 15E), EFS responses (F(1,7) = 10.431, p = 0.014; Figure 15F), and patch visits (F(1,4) = 6.8, p = 0.06; Figure 15G). Similarly, the animals showed a tendency to make more feeder well licks on amphetamine (F(1,4) = 5.714, p = 0.075; Figure 15H). These results suggest that the immediate effects of receiving reinforcement on responding are affected by amphetamine such that rats become more sensitive to reward delivery, whereas their sensitivity to reward omission is unaffected. Because systemic amphetamine increases levels of dopamine and other monoamines in the brain (Pontieri et al. 1995), these results further suggest that peak monoamine levels in some brain structure attenuates win-stay responding, whereas increasing the monoamine level at the nadir of its response (e.g. after reward omission) does not impact the loss signal involved in triggering lose-shift responding in our task. These results suggest that systemically elevated levels of dopamine selectively affect sensitivity to win outcomes while leaving intact motor functions, motivation, and responses to reward omission.

We next sought to determine which brain regions are affected by these manipulations. We previously found that win-stay and lose-shift responding dissociated among ventral and dorsolateral striatum (Skelin et al. 2014; Wong et al. 2016). We, therefore, inferred from the data above that the predominant effect of LHb lesions on responding in this task was the elevation of monoamines in the ventral striatum. We, therefore, compared the effects on this task of LHb lesions, systemic amphetamine injections, and local amphetamine microinfusions into the nucleus accumbens core from a previous study (Wong et al. 2016). Table 1 summarizes the data sets used for the comparison. Each of these treatments significantly decreased win-stay responses in the experimental group compared to their respective controls in an analogous manner (Figure 16A). However, none of the treatments affected lose-shift responding (Figure 16B). In addition, the response time of the animals in all three treatment groups decreased significantly (Figure 16C). Taken together, these results suggest that increased dopamine levels in the nucleus accumbens core either through direct microinfusion of dopamine agonists, systemic injection of amphetamine, or lesions of LHb, compromise an animal's ability to produce win-stay choices. Furthermore, this LHb circuitry does not appear to be involved in the production of lose-shift responses in a context lacking predictive cues.

Discussion

Emerging evidence suggests that LHb neurons play a critical role in response-switching based on recent outcomes (Matsumoto and Hikosaka 2007; Baker et al. 2015; Kawai et al. 2015; Mizumori and Baker 2017). Reward omissions stimulate the LHb (Salas et al. 2010; Kawai et al. 2015) and consequently cause a phasic pause in spiking activity in the dopaminergic neurons (Christoph et al. 1986; Ji and

Shepard 2007). Consequently, LHb has been identified as one of the primary brain region encoding negative prediction errors (Bromberg-Martin and Hikosaka 2011). However, little is known about the importance of these pauses and their influence on downstream structures and, ultimately, on an animal's decisions. A widely accepted theory on the behavioural significance of the dopamine pauses posits that LHb activation produces transiently low dopamine levels in the dorsal striatum which in turn inactivates D2 receptors and consequently causes the indirect pathway in the striatum to suppress the recent unrewarding actions, which thereby indirectly promotes shifting to alternate choices in the subsequent trials (Bromberg-Martin et al. 2010). Despite the detailed description of the LHb circuitry and its effects on downstream neurons, it is unclear (1) whether LHb drives choice adaptation based on wins or losses, or both, (2) whether such win-stay and lose-shift responses require predictable outcome cues, and (3) whether the LHb-dependent choice shifting occurs on a trial-to-trial basis or over several trials (Matsumoto and Hikosaka 2007; Kawai et al. 2015). In the present study, we used the unpredictable environment of the competitive choice task, which has the equal utility of both choice options, to reveal the memory systems involved in guiding choice over consecutive trials. We predicted that disruption of the ability of rats to generate reward prediction error signals either by the lesions of LHb or administration of amphetamine would decrease both lose-shift and win-stay responses. Extensive LHb damage (greater than 45%) did not affect the animal's ability to make lose-shift choices. Win-stay responses, however, were diminished, suggesting that win-stay and lose-shift strategies are controlled by different circuitry and that the rapid lose-shift observed in our task does not depend on LHb, or presumably depend on its control over dopaminergic neurons to produce reward prediction errors. Nonetheless, these results indicate that the LHb circuitry is involved in rapid win-stay adaptions when the environment lacks predictive reward cues. Systemic AMPH injection before the task decreased win-stay similarly while leaving lose-shift intact. Furthermore, our comparison of studies revealed that this behavioural pattern was also similar to localized microinfusions of AMPH into the nucleus accumbens core (Figure 16). All three manipulations (LHb lesions, systemic AMPH, localized AMPH) are expected to increase dopamine in the ventral striatum (Sugam et al. 2012). Together, these results suggest that above-normal levels of dopamine in the accumbens either interferes with the ability of this structure to generate win-stay responses or reduces the influence of accumbens on behavioural control of choice.

The LHb lesioned rats showed hyperactivity as compared to controls on several measures (Figure 14). The experimental group performed more trials within the same length of the test session, had a reduced response time, and showed more locomotion within the box as measured by distinct reward well and nose-poke entries (patch visits, Figure 14C). The effect of LHb lesions on motor activity is likely due to disruption of its tonic inhibition of dopaminergic neurons and consequent increases in baseline dopamine levels in downstream structures such as the nucleus accumbens. Indeed, lesions of the fasciculus retroflexus that conveys the habenulointerpeduncular tract have been shown to markedly increase dopamine levels in the cortex, nucleus accumbens, and striatum (Nishikawa et al. 1986). And we have previously reported that increasing tonic levels of dopamine either through systemic amphetamine injections or direct microinfusions of amphetamine into the nucleus accumbens core increases the number of trials completed in a session and decreases response time (Wong et al. 2016). Together, these findings suggest that the motoric hyperactivity induced by the habenula lesions is perhaps due to the absence of tonic inhibition of dopamine neurons. Furthermore, these motor effects validate the quality of the LHb lesions in the present experiment as being sufficiently large to make inferences on decision-making effects.

Functional Magnetic Resonance Imaging (fMRI) studies in humans (Salas et al. 2010) and electrophysiological studies in monkeys (Kawai et al. 2015) show that the LHb is sensitive to reward omissions and that it specifically encodes negative reward prediction error signals. This habenular activation is most likely used as a teaching signal by downstream structures to modify choices. For example, in a reversal learning task (Kawai et al. 2015) the probability of response-shifting increases after more than 2 consecutive reward omission outcomes. A key difference in our task, however, is that we are investigating the probability of shifting in the immediately next trial. We have previously reported that this shift tendency decays over 8 seconds and therefore does not accumulate over multiple trials (Chapter 3). This raises the possibility that a non-LHb dependent memory system, most likely involving the lateral striatum (Skelin et al. 2014) driving choice shifts to reward omission when the context has no predictable cues.

Chapter five

Lesions of Ventrolateral Striatum Eliminates Lose-shift but not Win-stay Behaviour in Rats[4]

Abstract

Animals tend to repeat actions that are associated with reward delivery, whereas they tend to shift responses to alternate choices following reward omission. These so-called win-stay and lose-shift responses are employed by a wide range of animals in a variety of decision-making scenarios, and these proximal effects of reinforcement often overshadow optimal actions based on the utility of choice options. These response strategies depend on dissociated regions of the striatum. Specifically, lose-shift responding is impaired by extensive excitotoxic lesions of the lateral striatum. Here we used focal lesions to assess whether dorsal and ventral regions of the lateral striatum contribute differently to this effect. We found that damage to ventrolateral striatum reduced lose-shift responding without impairing win-stay, motoric, or motivational aspects of behaviour in the task, whereas lesions confined to the dorsolateral striatum significantly impaired the ability of rats to complete trials of the task. Moreover, lesions to the dorsomedial striatum had no effect on either lose-shift or win-stay responding. Together, these data suggest a novel role of the ventral portion of the lateral striatum in driving lose-shift decisions.

[4] Accepted in the Neurobiology of Learning and Memory as: Thapa, R. and Gruber, A.G. (2018). Lesions of Ventrolateral Striatum Eliminates Lose-shift but not Win-stay Behaviour in Rats. Student’s Contribution: Designing the experiment, performing surgeries and histology, data collection, analysis, and writing the manuscript.

Introduction

Decision-making studies in animals often show a considerable influence of proximal factors on choice. For example, rats tend to place disproportionate weight on recent reinforcement and employ a simple heuristic of shifting choice to alternatives after a loss (lose-shift) and repeating a choice after a win (win-stay; Evenden and Robbins 1984). Although the win-stay and lose-shift strategies could in principle derive from a reinforcement learning mechanism with a large learning rate (Sutton and Barto 1998), several lines of evidence strongly support our proposal that these depend on dissociated brain circuits. First, win-stay and lose-shift have vastly different temporal dynamics (Gruber and Thapa 2016). The probability of lose-shift decays monotonically with increasing time between the reinforcement to the next choice, whereas the probability of win-stay first increases and then decreases in a parabolic manner with respect to the inter-trial interval (ITI). Second, they are differently affected by administration of d-amphetamine (Wong et al. 2016). Third, lesions of striatal regions produce dissociated effects on win-stay and lose-shift. Lesions of nucleus accumbens core in the ventromedial striatum reduced wins-stay but not lose-shift (Gruber et al. 2017), whereas lesions to the dorsolateral striatum (DLS) reduced lose-shift but not win-stay (Skelin et al. 2014; Gruber et al. 2017). In sum, the learning and memory systems supporting the influence of proximal wins and losses on decision-making appear to be dissociated from each other, and from other forms of reinforcement learning based on trial-and-error experience (Gruber and Thapa 2016).

Learning from past outcomes has often been attributed to the striatum (Graybiel 2008). Prior data suggest competing contributions of the lateral and medial parts of the dorsal striatum to reward-based choice learning (Balleine and O'Doherty 2010; Stalnaker et al. 2010; Bradfield and Balleine 2013; Brigman et al. 2013). Specifically, lesion studies show a functional segregation within the dorsal striatum such that DLS encodes stimulus-response associations that are formed gradually and support habitual responses (Jog 1999; Yin et al. 2004; Devan et al. 2011) whereas dorsomedial striatum (DMS) encodes flexible action-outcome associations and drives rapid response adaptations to changing contingencies (Yin et al. 2005; McDonald et al. 2008). In this conceptual framework, we would predict DMS lesions to impair behavioural flexibility required to execute rapid lose-shift and win-stay adaptations, whereas DLS lesions should have little effect on rapid choice shifts from trial-to-trial. Nonetheless, we previously found that DMS lesions have a mild effect on lose-shift whereas DLS lesions cause profound deficit, and neither of the lesions affects win-stay (Skelin et al. 2014). Furthermore, lesions of the lateral striatum in our previous studies were extensive and covered the entire lateral column of the striatum including the ventrolateral striatum (VLS). Although the VLS has received less attention in the decision-making literature, it is thought to be important for action initiation (Cousins et al. 1999; dos Santos et al. 2007; Tsutsui-Kimura et al. 2017). Here we used bilateral excitotoxic lesions confined to the DLS, DMS, or VLS in order to clarify the contributions of these striatal subregions to lose-shift and win-stay responses in rats.

Methods

Subjects

Twenty-eight male Long-Evans rats (Charles River, QC, Canada) weighing 400-500 g at the start of behavioural training were used in the lesion experiment. Animals were pair-housed in transparent plastic cages in a vivarium with a 12-hr light/dark cycle (lights off at 7:00 pm). The animals were handled for at least 2 minutes/day for one week before training commenced. Behavioural training and testing were conducted during the light phase (between 8:00 to 17:00 hours). The animals were given access to water in individual cages for 1 hour at the end of behavioural testing each day. Animals had ad libitum access to water on weekends and during surgical recovery. The water restriction protocol maintained the body weight of animals at > 85% of pre-testing weight. All procedures were pre-approved by the University of Lethbridge Animal Welfare Committee in accordance with the Canada Council of Animal Care guidelines.

Competitive choice task (CCT)

The CCT apparatus and procedure used in the present study are described in detail in Gruber and Thapa (2016). In brief, behavioural testing was performed in operant boxes containing two liquid delivery feeder wells and a central nose-poke port that were separated by a barrier orthogonal to the wall (Figure 17A). Individual trials of the task began with the illumination of two cue lights mounted proximally to the nose-poke port and the inactivation of the overhead house light. Animals then had 15 seconds to commit a nose-poke into the central port, and subsequently locomote to one of the two possible feeder wells of which only one was baited in each trial. If the rat selected the rewarded feeder, it received a drop of 10% sucrose solution (Win). If the rat chose the non-rewarded well, the feeder was left empty (Lose) and the house light was illuminated. The computer selected a priori which feeder well to reward based on a ‘competitive’ algorithm that uses the well choice and reinforcement history of the rat to predict the choice in the current trial and minimize the number of rewarded trials (Lee et al. 2004). The optimal solution for the rat in the competitive choice task is to be as random as possible by distributing its choices equally between the two feeder wells.

Surgery

At the end of training, the rats were randomly assigned to one of four lesion groups: dorsolateral striatum (DLS, n = 7), ventrolateral striatum (VLS, n = 7), dorsomedial striatum (DMS, n = 7), or control (n = 7) and excitotoxic lesion surgeries were performed. 30 min prior incision, all rats received Buprenorphine (Alstoe Ltd., UK) to mitigate pain. The animals were anesthetized using 4% isoflurane gas (Benson Medical Industries Inc., Ontario, Canada) in oxygen flowing at 1.0 L/min and the surgical plane was maintained with 2% isoflurane throughout the surgery. The animals were mounted on a stereotaxic frame (Kopf Instruments, Tujunga, CA, USA) and a midline incision was made to expose the skull. Burr holes were drilled through the skull to allow lowering of infusion cannulas at the following coordinates from bregma [in mm (AP, ML, DV)]: DLS (0.7, 3.6, -5.0); VLS (1.8, 3.8, -7.2); and DMS (0.4, 2.6, -4.6). The bilateral lesions were achieved by microinfusion of 0.1 M quinolinic acid (30 mg/mL in Dimethyl Sulfoxide, Sigma-Aldrich Canada Co., Oakville, Ontario, Canada). A total volume of 0.7 µl of quinolinic acid was infused at the rate of 0.2 µl/ min in each site using a 30-gauge injection cannula attached to a 10 µl Hamilton syringe via polyethylene tubing (PE-50). The injection needle was left in place for 3 min following the injection to allow diffusion of the drug. The control animals received the same treatment as the lesion group except for the infusion of phosphate buffered saline (PBS) in place of the quinolinic acid. The scalp incision was then closed with sutures. Rats were given subcutaneous injections (0.02 mg/kg) of meloxicam (Boehringer Ingelheim, Germany) and monitored for 24 hr before returning them to their home cage. All animals recovered in their home cages (pair housed) for at least 5 days before resuming behavioural testing.

Histology

At the experimental endpoint, all rats received lethal injections of sodium pentobarbital (100 mg/kg i.p.) and were perfused with physiological saline and 4% paraformaldehyde (PFA). The brains were post-fixed for 24 hr in PFA and then transferred and stored in 30% sucrose and PBS with sodium azide (0.02%) for a minimum of 48 h before sectioning. The brains were sectioned in the coronal plane at 40 µm thickness using an SM2010R freezing microtome (−19°C, Leica, Germany). Every third section through the region of interest was wet-mounted on glass microscope slides coated with 1% gelatin and 0.2% chromalum, allowed to air dry for at least a week and then stained with cresyl violet. Images of sections were digitized using a NanoZoomer (Hamamatsu, Japan) and evaluated for lesion quality.

Data analysis

We analyzed the data with MATLAB (version R2013a; MathWorks, MA, USA), Microsoft Excel (version 2016), and SPSS (version 21.0; IBM, NY, USA). One-way analysis of variance (ANOVA) and mixed ANOVA was used to assess the significance of lesion on behavioural measures ($p < 0.05$). Where the main effects were statistically significant, a post hoc Tukey test was used.

Rats in some trials sampled both feeder wells, a behaviour we termed extraneous feeder approach (EFS). We excluded all trials with EFS when computing lose-shift and win-stay decisions as EFS affects subsequent choice (Gruber et al. 2017). Moreover, because the tendency to lose-shift decays over a short ITI duration (Gruber and Thapa 2016), we excluded trials that were further than 7 seconds apart. The number of trials represents the total number of complete trials within a session. Only sessions with more than 100 trials were included in the analysis. The calculation of the percent of rewarded trials (win %) represents the percentage of all complete trials in which the rat was reinforced with sucrose. We computed the response time to be the time taken to reach the feeder after exiting nose-poke port. The EFS ratio was calculated as total EFS number over the total number of operant trials. Infrared sensors in the feeder wells recorded the licks rat made in each trial which was summated to give the total number of licks in a session. However, the total number of licks depends on the total number of wins which in turn depends on the total number of trials completed. Animals in separate groups may complete a significantly different number of trials making the measure of total licks incomparable. In addition, the mean number of licks on rewarded trials should be comparable between the groups because the amount of sucrose delivered was calibrated to be the same in each trial and across the testing boxes. Therefore, a ratio of total licks over total wins (licks/win) was computed to detect any abnormality in licking behaviour such as perseverative licking or incomplete sucrose consumption in each trial.

Results

The excitotoxic lesions produced focal striatal damage (Figure 17B-D). Two rats from the DMS lesion group and one from VLS had severe damage extending to the DLS subregion and were therefore excluded from analysis.

Lesions of striatal subregions can produce complex motoric and motivational deficiencies that may interfere with task performance (Eagle et al. 1999). Compromising VLS function, for instance, can produce intense orofacial stereotypies and impair feeding (Delfs and Kelley 1990; Salamone et al. 1993), but not affect drinking or locomotion (Bakshi and Kelley 1991; Baker et al. 1998). We, therefore, analyzed several measures of motoric function and motivation to identify any treatment effects that could confound choice performance (Figure 18). There were no significant differences in the number of trials completed among the VLS, DMS, or control groups between sessions before or after surgery (Figure 18A; $F_{12,90}$ = 1.218, p = 0.283). The DLS group had insufficient data for this analysis because they required re-training (with no barriers) for two sessions after surgery. We, therefore, next computed mean values from data aggregated from sessions following re-training. The DLS group alone had significantly fewer trials than controls (Figure 18B; ANOVA main effect: $F_{3,21}$ = 15.2, p = 1.7E-5; Tukey post-hoc shown in Figure 18B). There was a statistically significant difference between the groups in the response time as determined by one-way ANOVA (Figure 18C; $F_{3,21}$ = 16.102, p = 1E-6). A Tukey post hoc test revealed that the response time compared to controls (2.40 ± 0.438 s) was significantly higher in the DLS group (5.7 ± 1.8 s, p = 1E-6).

We accessed motivation by measuring the vigour of the licking responses at the reward wells. DLS lesions had a significantly higher number of licks in the rewarded trials from that of controls (Figure 18D; p = 0.013). Although the DLS animals completed significantly fewer trials, roughly half the number of trials were rewarded in each group (Figure 18E; $F_{3,21}$ = 1.08, p = 0.381). In sum, there were no strong motoric or motivation effects evident in the DMS or VLS groups. The DLS animals, however, completed far fewer trials and were much slower than any of the other groups. This presents a possible confound in the computation of lose-shift responding that we address later.

The DLS receives input from the sensorimotor cortex and is critical for learning tasks with a serial order of behavioural elements (Yin 2010). The operant response in our task comprises of two distinct serial responses: first, to enter the centre lane to make a nose-poke response and second, to enter one of the two reward-well lanes to check for sucrose delivery (Figure 17A). In addition, the animals can also choose to make an extraneous feeder sampling (EFS) response in which they shuttle directly between the reward wells, although this response is never rewarded. Because the cumulative sum of EFS responses increases with the number of trials completed, we computed the EFS ratio as the number of EFS responses normalized by the number of operant responses (completed trials). There was a significant difference in the EFS ratio among the groups (Figure 18F; $F_{3,21} = 13.0$, $p = 1E\text{-}6$). The EFS ratio for the control group was 0.13 ± 0.05. A Tukey post hoc test revealed that the EFS ratio was significantly higher than this in the animals with lesion to the DLS (0.82 ± 0.38, $p = 1E\text{-}6$), but this was not the case for animals with lesions to the DMS (0.22 ± 0.16, $p = 0.923$) or the VLS (0.29 ± 0.12, $p = 0.640$). Consistent with the notion that DLS is important for tasks with the serial order, DLS lesioned animals completed significantly less operant trials (Figure 18B) and instead completed a disproportionate number of unrewarded EFS responses (Figure 18F). The high EFS ratio and the considerable amount of licking (Figure 18D) in the DLS lesioned animals suggest that the motivation in the DLS animals was intact and that the decrease in the number of operant response is most likely due to disruption of serial order memory or motoric coordination supported by the sensorimotor striatum.

The post-surgery analysis of lose-shift and win-stay responses excluded the DLS lesioned animals because these animals performed significantly slower than the other groups, which affects the prevalence of lose-shift and win-stay responses within animals (Gruber and Thapa 2016). Specifically, the distribution of ITI for the DLS group did not sufficiently overlap that of the other groups, and we, therefore, did not have sufficient matching of a known confounding variable for a proper between-group contrast. For the remaining groups, we found a significant main effect of session (Fig 3A; $F_{6,12} = 3.34$, $p = 0.005$) and a significant lesion*session interaction ($F_{12,90} = 2.59$, $p = 0.005$). Post hoc Tukey analysis revealed that the animals with a lesion to the VLS had significantly lower lose-shift than the controls (*p* = 3.36E-4), whereas the animals with a lesion to the DMS were comparable to controls (*p* = 0.547). There was a statistically significant difference between groups of lose-shift probability as determined by one-way ANOVA (Fig 3B; $F_{2,15}$ = 14.8, *p* = 2.78E-4). A Tukey post hoc test revealed that lose-shift was significantly reduced by VLS lesions (0.47 ± 0.09, p = 2.02E-4; VLS-Control contrast), but not DMS lesions (0.61 ± 0.06, p = 0.209; DMS-Control contrast), relative to the controls (0.69 ± 0.07). In addition, the probability of lose-shift in VLS lesions was near chance levels for all ITI values, whereas that of controls and DMS lesioned animals showed the typical decay with increasing ITI (Figure 19C). The probability of win-stay was comparable between the groups in each session (Figure 19D; $F_{12,90}$ = 2.149, *p* = 0.065) and in aggregate (Figure 19E; $F_{2,15}$ = 0.297, *p* = 0.747). The win-stay tendency was similar across the relevant range of ITI between the groups (Figure 19F). Overall, the results suggest that lesions to the VLS attenuates lose-shift but leaves win-stay intact.

Discussion

The monotonic decay of the probability of lose-shift with increasing inter-trial interval points to the existence of a decaying influence supporting the short-lived influence of loss on choice (Gruber and Thapa 2016). The present results suggest that VLS is part of a learning and memory system supporting the transient decaying influence. Focal lesions of VLS were sufficient to reduce the robust lose-shift tendency seen in controls to chance levels (0.50). The reduction is not attributable to any motoric or motivational deficiencies because VLS animals had a similar number of trials, response time, EFS ratio, and mean number of licks on rewarded trials to controls. Moreover, the VLS lesion selectively attenuated lose-shift response while leaving win-stay intact. The negligible effect of DMS lesions on both lose-shift and win-stay supports the notion that these responses are more of a reflexive choice strategy rather than a product of either model-based or goal-oriented decisions posited to depend on this region of the striatum (Doll et al. 2012). Animals with DLS lesions performed a considerable number of EFS responses and licking, suggesting normal motivation. However, they appeared to have difficulty recalling the serial task contingencies which led to a substantial number of reward omissions and subsequent extinction of operant behaviour. Overall, the current results suggest that reduction of lose-shift responses observed in our previous reports of lateral striatum lesions is most likely due to damage to the VLS (Skelin et al. 2014; Gruber et al. 2017). The exclusion of subjects with an insufficient number of trials in the prior studies perhaps eliminated data from the rats with extensive DLS damage leaving rats with partially intact DLS, but sufficiently damaged VLS, in the analyzed cohort.

The VLS is implicated in action initiation (Cousins et al. 1999; dos Santos et al. 2007; Tsutsui-Kimura et al. 2017). Recently, Tsutsui-Kimura and colleagues (2017) have shown that medium spiny neurons in the VLS are activated at the time of cue presentation, which is early within the inter-trial interval in their task. In contrast, neurons in the ventromedial striatum are inhibited during this period. This opposing temporal patterns of ventral striatum neuronal activity shows a gradual decay over a 5 second period after a lever press response, which corresponds to the temporal decay of lose-shift tendency we observe in our task (Gruber and Thapa 2016). Although there are no explicit cues in our task, the transient decaying neuronal activity of the VLS neurons reported by Tsutsui-Kimura and colleagues (2017) could potentially represent the transient decaying influence supporting the initiation of the lose-shift response. These recent discoveries in VLS function, our present data, and our previous finding that win-stay depends on the integrity of ventromedial striatum (Gruber et al. 2017), together suggest a novel functional segregation within the ventral striatum such that the lateral region promotes lose-shift whereas the medial part supports win-stay behaviour.

Chapter six

General Discussion

Reward-oriented behaviours in animals are initially flexible, but with repeated trials, the response gradually becomes habitual and inflexible (Packard and McGaugh 1996). For example, imagine a cross-shaped maze in which rats are trained to run from the south arm to the west arm to retrieve food. If the experimenter suddenly changes the starting location to the north arm, rats early in training will make a right turn and go directly to the west arm. After several days of training, however, rats will make a habitual left turn and end up at the east arm. In the competitive choice task, the rats are trained to make repeated choices against a competitive agent that detects habitual responses and punishes predictable decisions with reward omission. Therefore, the choices of a trained rat in the CCT are always based on flexible responding. Several studies implicate the dorsal striatum for a shift from flexible to habitual choices. More specifically, lesion studies suggest functional segregation within the dorsal striatum such that dorsolateral striatum (DLS) encodes stimulus-response associations that are formed gradually and support habitual responses (Jog 1999; Yin et al. 2004; Devan et al. 2011) whereas dorsomedial striatum (DMS) encodes flexible action-outcome associations and drives rapid response adaptations to changing contingencies (Yin et al. 2005; McDonald et al. 2008). In this conceptual framework, we would predict DMS lesions to impair behavioural flexibility required to execute rapid lose-shift and win-stay adaptations, whereas DLS lesions should have little effect on rapid choice shifts from trial-to-trial. Nonetheless, Skelin et al. (2014) found that DMS lesions have a mild effect on lose-shift whereas DLS lesions cause profound deficit, and neither of the lesions affects win-stay (Skelin et al. 2014). This finding contradicts the decision-making role attributed to DLS in the literature. The underlying aim of the experiments outlined in this thesis was to investigate how DLS supports flexible lose-shift behaviour.

A large body of evidence suggests that dopamine in the striatum is important for reinforcement-driven choice adaptation (Gerfen et al. 1990; Kravitz et al. 2010). Both mesolimbic and nigrostriatal dopamine neurons generate bursting activation when reinforcements are better than expected, and they pause their firing when rewards are worse than expected (Schultz et al. 1997; Roesch et al. 2007). Current theories of reinforcement learning suggest that these properties of dopamine neurons provide a reward prediction error signal used by the striatum and other efferents to compute the value (Houk et al. 1995). Therefore, our first experiment (data not included in this thesis) investigated whether dopamine signalling is involved in the rapid reinforcement-driven choice adaptation following reward omission. In particular, we expected that the pauses in dopamine neurons that transmit negative reward prediction error signalling to cause an inactivation of the D2 type dopamine receptors (D2DR) because the high binding affinity of these receptors renders them partially activated by baseline dopamine levels (Montague et al. 1996; Schultz 1998; Steiner and Tseng 2010). D2DRs are expressed primarily by striatal projection cells involved in the ‘indirect’ pathway through the basal ganglia that have an inhibitory effect on thalamocortical activity and motor responses. In contrast, D1 type dopamine receptors (D1DR) are primarily expressed by striatal projection cells involved in the ‘direct’ pathway through the basal ganglia that have a disinhibitory effect on thalamocortical activity and motor responses. D1DR have a relatively lower binding affinity for dopamine, so we expected them to have little impact on rapid, trial-by-trial response adaptation following reward omission losses. Hence, we made local microinfusions of D2DR receptor agonist quinpirole into DLS and DMS. And we found that indeed disrupting D2DR function in the DLS, but not in DMS, reduced lose-shift. However, since quinpirole also affects D3DR we were less confident of our results and its interpretation. In collaboration with other students in the lab, we also tested whether globally suppressing negative reward signalling through systemic d-amphetamine administration would attenuate lose-shift, and once again we saw a dose-dependent decrease in lose-shift in the experimental group (Wong et al. 2016). A widely accepted theory on the behavioural significance of the dopamine pauses posits that lateral habenula (LHb) activation produces transiently low dopamine levels in the dorsal striatum via substantia nigra (SNc) disinhibition which in turn inactivates D2 receptors and consequently causes the indirect pathway in the striatum to suppress the recent unrewarding actions, which thereby indirectly promotes shifting to alternate choices in the subsequent trials (Bromberg-Martin et al. 2010). Therefore, the specific hypothesis we tested in this thesis is that the lose-shift responding observed in the CCT is controlled by a circuit linking the LHb, SNc, and DLS.

We repeated the lesions of DLS with the same coordinates used in Skelin et al. (2014). However, we added a new lesion group that of nucleus accumbens core (NACc) in the VMS which has been implicated in the reward-oriented task (Ikemoto and Panksepp 1999). We replicated the effect of DLS lesion on lose-shift. In addition, we learned that NACc lesions selectively attenuates win-stay behaviour while leaving lose-shift intact. We made electrolytic lesions of the LHb to attenuate dopamine pauses and consequently reduce lose-shift response. We failed to find any effect on either lose-shift or win-stay. In another experiment, we microinfused amphetamine into the NACc and the DLS. We expected disruption of dopamine signalling in the DLS by amphetamine to reduce lose-shift. There was no effect of amphetamine infusion into the DLS and that into the NACc only showed a general trend of decreased win-stay with increasing dosage.

We reanalyzed the behavioural data to investigate other pertinent variables in the task we had not considered yet. When we plotted histograms of inter-trial intervals, we discovered that it was positively skewed (right-skewed). In the trials with the longest ITI, it is possible that the animal might forget the outcome of the last trial and the subsequent decision is not reinforcement based. Therefore, we removed the trials with the longest inter-trial intervals that fell on the tail end of the skewed histogram. With this exclusion, we observed a much higher lose-shift tendency. Further excluding ITI at the tail end, further improved the lose-shift until it became clear at the shortest ITI of around 3-5 seconds, the probability of lose-shift was above 80% for many rats. Our data showed a clear log-linear relationship between the probability of lose-shift and ITI such that there is a gradual decrease in lose-shift with increasing ITI and by around 7 seconds the probability of lose-shift was almost at chance level (50%). The finding invigorated a more detailed investigation into the behavioural properties of lose-shift, win-stay, and all conceivable variables within the task. This led to the important discoveries that the lose-shift tendency is constant since the beginning of training whereas win-stay gradually increases across session with training. These findings and others became the eneuro paper described in Chapter two. We also discovered that animals at times shuttled directly between the two feeders and they tend to lose-shift from the last feeder visited rather than the first feeder checked. This behaviour we call extraneous feeder approach (EFS) became subject for the second eneuro paper detailed in Chapter three.

When we reanalyzed the previous systemic amphetamine administration data with the new exclusions that removed the effect of longer ITI and EFS on lose-shift, we found that there was only a marginal effect of increased dopamine levels on lose-shift. Moreover, when we reanalyzed the histology of DLS lesions, we found that the animals with the deepest lateral lesions were the ones with the most profound lose-shift deficit. It is possible that our DLS lesions in the prior studies were extensive and covered both the dorsal and ventral subregions of the lateral striatum, and that the effect on lose-shift was based on the VLS lesions alone. We also reanalyzed the LHb lesions data using stereological volume estimates and only included rats with significant damage confined to the LHb (> 45% volume loss). The animals with sufficient LHb lesion showed a clear deficit in win-stay response but had lose-shift response intact. We speculated that the permanent LHb damage lifted the tonic inhibition of dopamine neurons and in turn increased the baseline dopamine level in the downstream NACc and produced the similar effect of reduced win-stay seen in NACc amphetamine microinfusions. Because win-stay increases with training and reaches a plateau of around 60-70% only after extended training, prior experiments with too few training sessions would miss any effect on win-stay. We proposed that previous systemic amphetamine injection study showed little effect on win-stay because the animals were not trained long enough for the win-stay to reach its high plateau. When we repeated the experiment on well-trained animals, win-stay was indeed attenuated, but lose-shift was not affected (Chapter three).

We also repeated the striatal subregion lesions. However, we divided the lateral striatum lesions into focal DLS and VLS lesion groups. In addition, we included DMS lesions to confirm that win-stay and lose-shift were not dependent on model-based decision-making processes. The results confirmed our prediction that VLS lesions alone can selectively attenuate lose-shift (Chapter five). Interestingly, animals with DLS lesions had difficulty completing enough trials. DLS is implicated in serial order memory (Yin 2010) and most likely the two-step serial ordered nature of the CCT was difficult to complete with damage to the sensorimotor striatum (DLS). Histological exclusions of rats with fewer trials in our past lateral striatum lesion studies likely excluded animals with extensive DLS damage and left animals with partial but functional DLS and damaged VLS in the analyzed cohort. DMS lesions did not affect either lose-shift or win-stay.

If the LHb-SNc-DLS circuit did not convey the loss information needed by VLS to execute lose-shift, where does the negative prediction error signal in the VLS originate from? Could the error signal to the VLS be coming from the cortex? Previously, we had made electrolytic focal lesions of different cortical areas (anterior cingulate cortex, prelimbic cortex, and orbitofrontal cortex) and tested the animals on the CCT to find no effect on lose-shift and win-stay. Only the lesions of the orbitofrontal cortex (OFC) had few animals with mixed effects on decision-making (unpublished data). Another student in the lab had followed up the experiment with chemical lesions of OFC subregions—namely, lateral orbitofrontal cortex (lOFC) and medial orbitofrontal cortex lesions (mOFC) and found the same inconclusive results (unpublished data). We reanalyzed the histology from the cortical electrolytic lesions and the second chemical lesions and found that in both cohorts, animals with lesions to more medial parts of OFC appeared to have a higher lose-shift whereas animals with the most lateral lesions had reduced lose-shift. The lateral orbitofrontal circuit has been implicated in animal's capacity to make appropriate switches in behavioural sets (Divac et al. 1967; Mishkin and Manning 1978; Alexander et al. 1986). And more recently there has been a proposal for functional compartmentalization within the lateral striatum such that the dorsal part is important for stimulus-response habit system and the ventral compartment with its lOFC connections is important for outcome-based decision-making (Gourley et al. 2013). Could the lOFC-VLS circuit be driving the lose-shift response? We followed this curiosity with another OFC subregion lesions of the lOFC and mOFC. We found that indeed the lesions to the mOFC increased lose-shift tendency whereas lOFC lesions had no effect (data not included in this thesis). VLS also receives major glutamatergic afferents from the insular cortex (IC) which is more lateral to the lOFC (Berendse et al. 1992). It is possible that the decrease in lose-shift seen in the lateral most electrolytic lesions of OFC affected IC, and that the IC-VLS circuit drives lose-shift. Another possibility we have yet to consider is that the shift we see after reward omission in CCT is outcome independent. In other words, perhaps it is a fundamental property of the basal ganglia action selection system that for the shortest ITIs a choice-shift is promoted regardless of win or loss in the preceding trial. Indeed, we do notice a higher win-shift tendency at earlier ITI (Chapter two). And VLS lesions tend to attenuate the higher win-shift tendency usually seen at earlier ITI in controls (Chapter five). With these observations, there are at least two experiments that need to be completed. First, investigation of the possibility of IC-VLS circuit driving lose-shift. Second, a collection of more samples of win-stay at earlier ITI using optogenetic self-stimulation of VTA as a reward instead of sucrose and thereby test the dependence of choice-shifts at earlier ITI on reinforcement outcome. Beyond that, future experiments should examine the following important questions raised by the results presented in this thesis:

1. *Dorsal versus ventral striatal function:* Our data suggest a novel functional segregation within the ventral striatum such that the lateral part drives lose-shift (or simply choice-shift) and the medial part drives win-stay. Both choice heuristics are reflexive and based on proximal data. The dorsal striatum has been proposed to be important for goal-oriented trial-and-error learning. Could we segregate the striatum functionally along the vertical axis such that the ventral parts influence choice based on proximal variables whereas the dorsal part drives choice based on the integration of reinforcement history? How do the ventral reflexive decision tendencies interact with the dorsal reinforcement-based learning? Do the dorsal and ventral striatum compete for behavioural control?
2. *Nature of VLS-VMS interaction:* Neural activity in VMS is suppressed while that in VLS is high at an earlier time point within the ITI (Tsutsui-Kimura et al. 2017). Is win-stay actively suppressed when lose-shift is promoted at earlier ITI? Which inputs to VMS suppress win-stay?

3. *Transfer of value:* How does the initial reflexive win-stay memory system transfer value to the long-term trial-and-error learning of the dorsal striatum?

Conclusion

Here we investigate how decisions are affected by recent wins and loss, inter-trial interval, and level of contextual uncertainty. We report that the tendency to shift to alternatives after a loss and to repeat a response after a win are dissociable from trial-and-error based reinforcement learning system and from each other. Furthermore, uncertainty driven exploratory EFS is dissociable from the outcome-driven behaviours. We observed that the probability of lose-shift decays monotonically with increasing ITI, whereas the probability of win-stay first increases and then decreases in a parabolic manner with respect to ITI. EFS responses steadily decrease with training but are never diminished despite lack of positive reinforcement and it can increase with the introduction of novel contextual elements. Lesions of ventrolateral striatum selectively attenuate lose-shift but not win-stay whereas lesions of ventromedial striatum decrease win-stay but not lose-shift. Disrupting dopamine signalling either systemically or with local microinfusions diminishes win-stay but not lose-shift. Together, the present data suggest that the ventral striatum is important for mediating the influence of proximal factors on choice. Specifically, we provide evidence for a novel functional segregation within the ventral striatum such that the lateral side drives lose-shift and the medial side guides win-stay behaviour.

A fundamental goal of the enterprise of behavioural neuroscience is to map animal behaviour to brain structure and circuitry. Here we have shown that the ventral striatum is critical for the production of adaptive choices driven by recent reinforcement outcome. How dopaminergic reward prediction error teaching signal translates to action selection in the downstream structures is an important question to answer (Bromberg-Martin et al. 2010). With the use of multiple lesion studies and neuropharmacology, we have established a clear link between the phasic dopaminergic input to the VMS and win-stay behaviour (Wong et al. 2016; Gruber et al. 2017). The drive to repeat a response is at the heart of how animals come to be shaped on a new behavioural pattern in normal circumstances and how compulsive action might develop in the case of drug addiction. Therefore, our findings provide a promising starting point to investigate how the action of dopamine on the VMS neurons leads to a drive for response repetition. Lastly, the novel role of VLS in driving lose-shift response points to the possibility of either a hitherto undescribed error detection system separate from that of the dopaminergic signals that the brain uses to implement choice shifts after reward omission or a reflexive choice shift mechanism that the basal ganglia employ when making rapid successive decisions. In either case, future experiments invigorated by our present findings will inform us about a fundamental property of the basal ganglia in action selection.

References

Alexander GE, DeLong MR, Strick PL. 1986. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual review of neuroscience 9: 357-381.

Alexander WH, Brown JW. 2011. Medial prefrontal cortex as an action-outcome predictor. Nature neuroscience 14: 1338-1344.

Altmann EM, Gray WD. 2002. Forgetting to remember: The functional relationship of decay and interference. Psychological science 13: 27-33.

Amodeo DA, Jones JH, Sweeney JA, Ragozzino ME. 2012. Differences in BTBR T+ tf/J and C57BL/6J mice on probabilistic reversal learning and stereotyped behaviors. Behavioural brain research 227: 64-72.

Amsel A. 1958. The role of frustrative nonreward in noncontinuous reward situations. Psychological bulletin 55: 102.

Aparicio CF. 2001. Overmatching in rats: The barrier choice paradigm. Journal of the Experimental Analysis of Behavior 75: 93-106.

Aron AR. 2011. From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. Biological psychiatry 69: e55-68.

Baker DA, Specio SE, Tran-Nguyen LT, Neisewander JL. 1998. Amphetamine infused into the ventrolateral striatum produces oral stereotypies and conditioned place preference. Pharmacol Biochem Be 61: 107-111.

Baker PM, Jhou T, Li B, Matsumoto M, Mizumori SJ, Stephenson-Jones M, Vicentic A. 2016. The Lateral Habenula Circuitry: Reward Processing and Cognitive Control. The Journal of neuroscience : the official journal of the Society for Neuroscience 36: 11482-11488.

Baker PM, Oh SE, Kidder KS, Mizumori SJ. 2015. Ongoing behavioral state information signaled in the lateral habenula guides choice flexibility in freely moving rats. Front Behav Neurosci 9: 295.

Bakshi V, Kelley AE. 1991. Dopaminergic regulation of feeding behavior: I. Differential effects of haloperidol microinfusion into three striatal subregions. Psychobiology 19: 223-232.

Balcita-Pedicino JJ, Omelchenko N, Bell R, Sesack SR. 2011. The inhibitory influence of the lateral habenula on midbrain dopamine cells: ultrastructural evidence for indirect mediation via the rostromedial mesopontine tegmental nucleus. The Journal of comparative neurology 519: 1143-1164.

Balleine BW, O'Doherty JP. 2010. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology 35: 48-69.

Bari A, Robbins TW. 2013. Inhibition and impulsivity: behavioral and neural basis of response control. Progress in neurobiology 108: 44-79.

Barraclough DJ, Conroy ML, Lee D. 2004. Prefrontal cortex and decision making in a mixed-strategy game. Nature neuroscience 7: 404-410.

Baum WM. 1974. On two types of deviation from the matching law: Bias and undermatching. Journal of the experimental analysis of behavior 22: 231-242.

Berendse HW, Graaf YGD, Groenewegen HJ. 1992. Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. Journal of Comparative Neurology 316: 314-347.

Blaiss CA, Janak PH. 2009. The nucleus accumbens core and shell are critical for the expression, but not the consolidation, of Pavlovian conditioned approach. Behavioural brain research 200: 22-32.

Boakes RA. 1977. Performance on learning to associate a stimulus with positive reinforcement. Operant-Pavlovian interactions: 67-97.

Bradfield LA, Balleine BW. 2013. Hierarchical and binary associations compete for behavioral control during instrumental biconditional discrimination. Journal of Experimental Psychology: Animal Behavior Processes 39: 2.

Breland K, Breland M. 1961. The misbehavior of organisms. American psychologist 16: 681-684.

Brigman JL, Daut RA, Wright T, Gunduz-Cinar O, Graybeal C, Davis MI, Jiang Z, Saksida LM, Jinde S, Pease M. 2013. GluN2B in corticostriatal circuits governs choice learning and choice shifting. Nature neuroscience 16: 1101-1110.

Brinschwitz K, Dittgen A, Madai V, Lommel R, Geisler S, Veh R. 2010. Glutamatergic axons from the lateral habenula mainly terminate on GABAergic neurons of the ventral midbrain. Neuroscience 168: 463-476.

Bromberg-Martin ES, Hikosaka O. 2011. Lateral habenula neurons signal errors in the prediction of reward information. Nature neuroscience 14: 1209-1216.

Bromberg-Martin ES, Matsumoto M, Hikosaka O. 2010. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron 68: 815-834.

Brown PL, Palacorolla H, Brady D, Riegger K, Elmer GI, Shepard PD. 2017. Habenula-induced inhibition of midbrain dopamine neurons is diminished by lesions of the rostromedial tegmental nucleus. Journal of Neuroscience 37: 217-225.

Carli M, Robbins TW, Evenden JL, Everitt BJ. 1983. Effects of lesions to ascending noradrenergic neurones on performance of a 5-choice serial reaction task in rats; implications for theories of dorsal noradrenergic bundle function based on selective attention and arousal. Behavioural brain research 9: 361-380.

Christoph GR, Leonzio RJ, Wilcox KS. 1986. Stimulation of the lateral habenula inhibits dopamine-containing neurons in the substantia nigra and ventral tegmental area of the rat. The Journal of neuroscience 6: 613-619.

Cohen JY, Amoroso MW, Uchida N. 2015. Serotonergic neurons signal reward and punishment on multiple timescales. Elife 4: e06346.

Cousins MS, Trevitt J, Atherton A, Salamone JD. 1999. Different behavioral functions of dopamine in the nucleus accumbens and ventrolateral striatum: a microdialysis and behavioral investigation. Neuroscience 91: 925-934.

Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature neuroscience 8: 1704-1711.

Daw ND, O'doherty JP, Dayan P, Seymour B, Dolan RJ. 2006. Cortical substrates for exploratory decisions in humans. Nature 441: 876-879.

Day JJ, Jones JL, Carelli RM. 2011. Nucleus accumbens neurons encode predicted and ongoing reward costs in rats. European journal of neuroscience 33: 308-321.

Dayan P, Niv Y, Seymour B, Daw ND. 2006. The misbehavior of value and the discipline of the will. Neural Netw 19: 1153-1160.

Delfs JM, Kelley AE. 1990. The role of D1 and D2 dopamine receptors in oral stereotypy induced by dopaminergic stimulation of the ventrolateral striatum. Neuroscience 39: 59-67.

Derenne A, Flannery KA. 2007. Within-session changes in the preratio pause on fixed-ratio schedules of reinforcement. The Behavior Analyst Today 8: 175-186.

Devan BD, Hong NS, McDonald RJ. 2011. Parallel associative processing in the dorsal striatum: segregation of stimulus-response and cognitive control subregions. Neurobiology of learning and memory 96: 95-120.

Divac I, Rosvold HE, Szwarcbart MK. 1967. Behavioral effects of selective ablation of the caudate nucleus. Journal of comparative and physiological psychology 63: 184.

Dolan RJ, Dayan P. 2013. Goals and habits in the brain. Neuron 80: 312-325.

Doll BB, Simon DA, Daw ND. 2012. The ubiquity of model-based reinforcement learning. Curr Opin Neurobiol 22: 1075-1081.

dos Santos LM, Ferro MM, Mota-Ortiz SR, Baldo MV, da Cunha C, Canteras NS. 2007. Effects of ventrolateral striatal inactivation on predatory hunting. Physiol Behav 90: 669-673.

du Hoffmann J, Nicola SM. 2014. Dopamine invigorates reward seeking by promoting cue-evoked excitation in the nucleus accumbens. The Journal of neuroscience : the official journal of the Society for Neuroscience 34: 14349-14364.

Eagle DM, Humby T, Dunnett SB, Robbins TW. 1999. Effects of regional striatal lesions on motor, motivational, and executive aspects of progressive-ratio performance in rats. Behav Neurosci 113: 718-731.

Estes WK, Skinner BF. 1941. Some quantitative properties of anxiety. Journal of Experimental Psychology 29: 390.

Euston DR, Gruber AJ, McNaughton BL. 2012. The role of medial prefrontal cortex in memory and decision making. Neuron 76: 1057-1070.

Evenden J, Robbins T. 1984. Win-stay behaviour in the rat. The Quarterly Journal of Experimental Psychology 36: 1-26.

Evenden JL. 1999. Varieties of impulsivity. Psychopharmacology (Berl) 146: 348-361.

Farwell BJ, Ayres JJ. 1979. Stimulus-reinforcer and response-reinforcer relations in the control of conditioned appetitive headpoking (“goal tracking”) in rats. Learning and Motivation 10: 295-312.

Felton M, Lyon DO. 1966. The post-reinforcement pause. Journal of the Experimental Analysis of Behavior 9: 131-134.

Ferster CB, Skinner BF. 1957. Schedules of reinforcement.

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. 2007. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proceedings of the National Academy of Sciences 104: 16311-16316.

Gan JO, Walton ME, Phillips PE. 2010. Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nature neuroscience 13: 25-27.

Gerfen CR, Engber TM, Mahan LC, Susel Z, Chase TN, Monsma FJ, Jr., Sibley DR. 1990. D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. Science 250: 1429-1432.

Glimcher PW. 2011. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proceedings of the National Academy of Sciences 108: 15647-15654.

Goncalves L, Sego C, Metzger M. 2012. Differential projections from the lateral habenula to the rostromedial tegmental nucleus and ventral tegmental area in the rat. The Journal of comparative neurology 520: 1278-1300.

Gourley SL, Olevska A, Zimmermann KS, Ressler KJ, DiLeone RJ, Taylor JR. 2013. The orbitofrontal cortex regulates outcome-based decision-making via the lateral striatum. European Journal of Neuroscience 38: 2382-2388.

Graybiel AM. 1998. The basal ganglia and chunking of action repertoires. Neurobiology of learning and memory 70: 119-136.

Graybiel AM. 2008. Habits, rituals, and the evaluative brain. Annu Rev Neurosci 31: 359-387.

Gruber AJ, Calhoon GG, Shusterman I, Schoenbaum G, Roesch MR, O'Donnell P. 2010. More is less: a disinhibited prefrontal cortex impairs cognitive flexibility. The Journal of neuroscience : the official journal of the Society for Neuroscience 30: 17102-17110.

Gruber AJ, McDonald RJ. 2012. Context, emotion, and the strategic pursuit of goals: interactions among multiple brain systems controlling motivated behavior. Frontiers in behavioral neuroscience 6: 50-50.

Gruber AJ, Thapa R. 2016. The memory trace supporting lose-shift responding decays rapidly after reward omission and is distinct from other learning mechanisms in rats. eneuro 3: ENEURO. 0167-0116.2016.

Gruber AJ, Thapa R, Randolph SH. 2017. Feeder approach between trials is increased by uncertainty and affects subsequent choices. eNeuro: ENEURO. 0437-0417.2017.

Herrnstein RJ. 1961. RELATIVE AND ABSOLUTE STRENGTH OF RESPONSE AS A FUNCTION OF FREQUENCY OF REINFORCEMENT1, 2. Journal of the experimental analysis of behavior 4: 267-272.

Holland PC. 2004. Relations between Pavlovian-instrumental transfer and reinforcer devaluation. Journal of Experimental Psychology: Animal Behavior Processes 30: 104.

Hori Y, Minamimoto T, Kimura M. 2009. Neuronal encoding of reward value and direction of actions in the primate putamen. Journal of neurophysiology 102: 3530-3543.

Houk JC, Davis JL, Beiser DG. 1995. Models of information processing in the basal ganglia. MIT press.

Hu J, Li X, Yin J. 2010. Learning and Decision Making in Human During a Game of Matching Pennies. JDCTA 4: 100-108.

Ikemoto S, Panksepp J. 1999. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. Brain Research Reviews 31: 6-41.

Ito M, Doya K. 2009. Validation of decision-making models and analysis of decision variables in the rat basal ganglia. The Journal of neuroscience : the official journal of the Society for Neuroscience 29: 9861-9874.

Ji H, Shepard PD. 2007. Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a GABA(A) receptor-mediated mechanism. The Journal of neuroscience : the official journal of the Society for Neuroscience 27: 6923-6930.

Jog MS. 1999. Building Neural Representations of Habits. Science 286: 1745-1749.

Kahneman D, Tversky A. 1979. Prospect theory: An analysis of decision under risk. Econometrica: Journal of the econometric society: 263-291.

Kakade S, Dayan P. 2002. Dopamine: generalization and bonuses. Neural Netw 15: 549-559.

Kawai T, Yamada H, Sato N, Takada M, Matsumoto M. 2015. Roles of the Lateral Habenula and Anterior Cingulate Cortex in Negative Outcome Monitoring and Behavioral Adjustment in Nonhuman Primates. Neuron 88: 792-804.

Komischke B, Giurfa M, Lachnit H, Malun D. 2002. Successive olfactory reversal learning in honeybees. Learning & memory 9: 122-129.

Kravitz AV, Freeze BS, Parker PR, Kay K, Thwin MT, Deisseroth K, Kreitzer AC. 2010. Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. Nature 466: 622-626.

Kravitz AV, Tye LD, Kreitzer AC. 2012. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nature neuroscience 15: 816-818.

Lee D, Conroy ML, McGreevy BP, Barraclough DJ. 2004. Reinforcement learning and decision making in monkeys during a competitive game. Cognitive Brain Research 22: 45-58.

Lesaint F, Sigaud O, Flagel SB, Robinson TE, Khamassi M. 2014. Modelling individual differences in the form of Pavlovian conditioned approach responses: a dual learning systems approach with factored representations. PLoS Comput Biol 10: e1003466.

Matsumoto M, Hikosaka O. 2007. Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447: 1111-1115.

Matsumoto M, Matsumoto K, Abe H, Tanaka K. 2007. Medial prefrontal cell activity signaling prediction errors of action values. Nature neuroscience 10: 647-656.

McDonald RJ, King AL, Foong N, Rizos Z, Hong NS. 2008. Neurotoxic lesions of the medial prefrontal cortex or medial striatum impair multiple-location place learning in the water task: evidence for neural structures with complementary roles in behavioural flexibility. Experimental brain research 187: 419-427.

McDonald RJ, White NM. 1993. A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. Behavioral neuroscience 107: 3.

-. 1995. Information acquired by the hippocampus interferes with acquisition of the amygdala-based conditioned-cue preference in the rat. Hippocampus 5: 189-197.

McGeorge A, Faull R. 1989. The organization of the projection from the cerebral cortex to the striatum in the rat. Neuroscience 29: 503-537.

Means LW, Fernandez TJ. 1992. Daily glucose injections facilitate performance of a win-stay water-escape working memory task in mice. Behavioral neuroscience 106: 345.

Mishkin M, Manning FJ. 1978. Non-spatial memory after selective prefrontal lesions in monkeys. Brain research 143: 313-323.

Mishkin M, Prockop ES, Rosvold HE. 1962. One-trial object-discrimination learning in monkeys with frontal lesions. Journal of comparative and physiological psychology 55: 178.

Mizumori SJ, Baker PM. 2017. The Lateral Habenula and Adaptive Behaviors. Trends in Neurosciences 40: 481-493.

Mizumori SJ, Channon V, Rosenzweig MR, Bennett EL. 1987. Short-and long-term components of working memory in the rat. Behavioral neuroscience 101: 782.

Moeller FG, Barratt ES, Dougherty DM, Schmitz JM, Swann AC. 2001. Psychiatric aspects of impulsivity. Am J Psychiatry 158: 1783-1793.

Montague PR, Dayan P, Sejnowski TJ. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. The Journal of neuroscience 16: 1936-1947.

Morrison SE, Bamkole MA, Nicola SM. 2015. Sign tracking, but not goal tracking, is resistant to outcome devaluation. Frontiers in neuroscience 9.

Mowrer O. 1947. On the dual nature of learning—a re-interpretation of" conditioning" and" problem-solving.". Harvard educational review.

Murphy CA, DiCamillo AM, Haun F, Murray M. 1996. Lesion of the habenular efferent pathway produces anxiety and locomotor hyperactivity in rats: a comparison of the effects of neonatal and adult lesions. Behavioural brain research 81: 43-52.

Nicola SM. 2010. The flexible approach hypothesis: unification of effort and cue-responding hypotheses for the role of nucleus accumbens dopamine in the activation of reward-seeking behavior. The Journal of neuroscience : the official journal of the Society for Neuroscience 30: 16585-16600.

Nishikawa T, Fage D, Scatton B. 1986. Evidence for, and nature of, the tonic inhibitory influence of habenulointerpeduncular pathways upon cerebral dopaminergic transmission in the rat. Brain research 373: 324-336.

Olton DS, Walker JA, Gage FH. 1978. Hippocampal connections and spatial discrimination. Brain research 139: 295-308.

Packard MG, McGaugh JL. 1996. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiology of learning and memory 65: 65-72.

Parkinson JA, Olmstead MC, Burns LH, Robbins TW, Everitt BJ. 1999. Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity byd-amphetamine. The Journal of Neuroscience 19: 2401-2411.

Paton JJ, Belova MA, Morrison SE, Salzman CD. 2006. The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature 439: 865-870.

Pennartz CM, Berke JD, Graybiel AM, Ito R, Lansink CS, Van Der Meer M, Redish AD, Smith KS, Voorn P. 2009. Corticostriatal interactions during learning, memory processing, and decision making. Journal of Neuroscience 29: 12831-12838.

Pontieri F, Tanda G, Di Chiara G. 1995. Intravenous cocaine, morphine, and amphetamine preferentially increase extracellular dopamine in the" shell" as compared with the" core" of the rat nucleus accumbens. Proceedings of the National Academy of Sciences 92: 12304-12308.

Quinn JJ, Pittenger C, Lee AS, Pierson JL, Taylor JR. 2013. Striatum-dependent habits are insensitive to both increases and decreases in reinforcer value in mice. European Journal of Neuroscience 37: 1012-1021.

Rayburn-Reeves RM, Laude JR, Zentall TR. 2013. Pigeons show near-optimal win-stay/lose-shift performance on a simultaneous-discrimination, midsession reversal task with short intertrial intervals. Behavioural processes 92: 65-70.

Rescorla RA, Solomon RL. 1967. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychological review 74: 151.

Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 2: 64-99.

Reynolds B, De Wit H, Richards JB. 2002. Delay of gratification and delay discounting in rats. Behavioural Processes 59: 157-168.

Robinson TE, Flagel SB. 2009. Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences. Biological psychiatry 65: 869-873.

Roesch MR, Calu DJ, Schoenbaum G. 2007. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nature neuroscience 10: 1615-1624.

Ross RR. 1964. Positive and negative partial-reinforcement extinction effects carried through continuous reinforcement, changed motivation, and changed response. Journal of Experimental Psychology 68: 492.

Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW. 2009. Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. The Journal of Neuroscience 29: 15104-15114.

Salamone JD, Mahan K, Rogers S. 1993. Ventrolateral striatal dopamine depletions impair feeding and food handling in rats. Pharmacol Biochem Behav 44: 605-610.

Salas R, Baldwin P, de Biasi M, Montague PR. 2010. BOLD Responses to Negative Reward Prediction Errors in Human Habenula. Frontiers in human neuroscience 4: 36.

Samejima K, Ueda Y, Doya K, Kimura M. 2005. Representation of action-specific reward values in the striatum. Science 310: 1337-1340.

Schmitz C, Hof PR. 2005. Design-based stereology in neuroscience. Neuroscience 130: 813-831.

Schönberg T, Daw ND, Joel D, O'Doherty JP. 2007. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. Journal of Neuroscience 27: 12860-12867.

Schultz W. 1998. Predictive reward signal of dopamine neurons. Journal of neurophysiology 80: 1-27.

Schultz W. 2013. Updating dopamine reward signals. Curr Opin Neurobiol 23: 229-238.

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. Science 275: 1593-1599.

Schusterman RJ. 1962. Transfer effects of successive discrimination-reversal training in chimpanzees. Science 137: 422-423.

Shimp CP. 2014. What Means in Molecular, Molar, and Unified Analyses. International Journal of Comparative Psychology 27.

Skelin I, Hakstol R, VanOyen J, Mudiayi D, Molina LA, Holec V, Hong NS, Euston DR, McDonald RJ, Gruber AJ. 2014. Lesions of dorsal striatum eliminate lose-switch responding but not mixed-response strategies in rats. European Journal of Neuroscience 39: 1655-1663.

Solomon RL, Corbit JD. 1974. An opponent-process theory of motivation. I. Temporal dynamics of affect. Psychol Rev 81: 119-145.

Staddon J, Motheral S. 1978. On matching and maximizing in operant choice experiments. Psychological Review 85: 436.

Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G. 2010. Neural correlates of stimulus–response and response–outcome associations in dorsolateral versus dorsomedial striatum. Frontiers in integrative neuroscience 4.

Steiner H, Tseng KY. 2010. Handbook of basal ganglia structure and function: a decade of progress. Academic Press.

Stote DL, Fanselow MS. 2004. NMDA receptor modulation of incidental learning in Pavlovian context conditioning. Behavioral neuroscience 118: 253.

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

Sugrue LP, Corrado GS, Newsome WT. 2004. Matching behavior and the representation of value in the parietal cortex. Science 304: 1782-1787.

Sul JH, Jo S, Lee D, Jung MW. 2011. Role of rodent secondary motor cortex in value-based action selection. Nature neuroscience 14: 1202-1208.

Sutton RS, Barto AG. 1998. Reinforcement learning: An introduction. MIT press Cambridge.

Tervo DG, Proskurin M, Manakov M, Kabra M, Vollmer A, Branson K, Karpova AY. 2014. Behavioral variability through stochastic choice and its gating by anterior cingulate cortex. Cell 159: 21-32.

Thorndike EL. 1911. Edward Lee Thorndike. Anim Intell 1874: 1949.

Thorndike EL. 1927. The law of effect. The American Journal of Psychology 39: 212-222.

Thornton EW, Davies C. 1991. A water-maze discrimination learning deficit in the rat following lesion of the habenula. Physiol Behav 49: 819-822.

Tian J, Uchida N. 2015. Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. Neuron 87: 1304-1316.

Tinbergen N. 1963. On aims and methods of ethology. Ethology 20: 410-433.

Tsutsui-Kimura I, Natsubori A, Mori M, Kobayashi K, Drew MR, de Kerchove d'Exaerde A, Mimura M, Tanaka KF. 2017. Distinct Roles of Ventromedial versus Ventrolateral Striatal Medium Spiny Neurons in Reward-Oriented Behavior. Current Biology 27: 3042-3048. e3044.

van der Meer M, Kurth-Nelson Z, Redish AD. 2012. Information processing in decision-making systems. The Neuroscientist 18: 342-359.

Vertes RP. 2004. Differential projections of the infralimbic and prelimbic cortex in the rat. Synapse 51: 32-58.

Viswanath H, Carter AQ, Baldwin PR, Molfese DL, Salas R. 2013. The medial habenula: still neglected. Frontiers in human neuroscience 7.

Voorn P, Vanderschuren LJ, Groenewegen HJ, Robbins TW, Pennartz C. 2004. Putting a spin on the dorsal–ventral divide of the striatum. Trends in neurosciences 27: 468-474.

Wang Z, Xu B, Zhou H-J. 2014. Social cycling and conditional responses in the Rock-Paper-Scissors game. arXiv preprint arXiv:14045199.

Watkins CJ, Dayan P. 1992. Q-learning. Machine learning 8: 279-292.

Williams BA. 1991. Choice as a function of local versus molar reinforcement contingencies. Journal of the experimental analysis of behavior 56: 455-473.

Williams DR, Williams H. 1969. AUTO-MAINTENANCE IN THE PIGEON: SUSTAINED PECKING DESPITE CONTINGENT NON-REINFORCEMENT2. Journal of the experimental analysis of behavior 12: 511-520.

Wong SA, Randolph SH, Ivan VE, Gruber AJ. 2017. Acute Δ-9-tetrahydrocannabinol administration in female rats attenuates immediate responses following losses but not multi-trial reinforcement learning from wins. Behavioural brain research 335: 136-144.

Wong SA, Thapa R, Badenhorst CA, Briggs AR, Sawada JA, Gruber AJ. 2016. Opposing effects of acute and chronic d-amphetamine on decision-making in rats. Neuroscience.

Yin HH. 2010. The sensorimotor striatum is necessary for serial order learning. Journal of Neuroscience 30: 14719-14723.

Yin HH, Knowlton BJ. 2004. Contributions of striatal subregions to place and response learning. Learning & memory 11: 459-463.

Yin HH, Knowlton BJ, Balleine BW. 2004. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. The European journal of neuroscience 19: 181-189.

Yin HH, Ostlund SB, Knowlton BJ, Balleine BW. 2005. The role of the dorsomedial striatum in instrumental conditioning. The European journal of neuroscience 22: 513-523.

TABLES

Table 1 is published in Gruber & Thapa (2016).

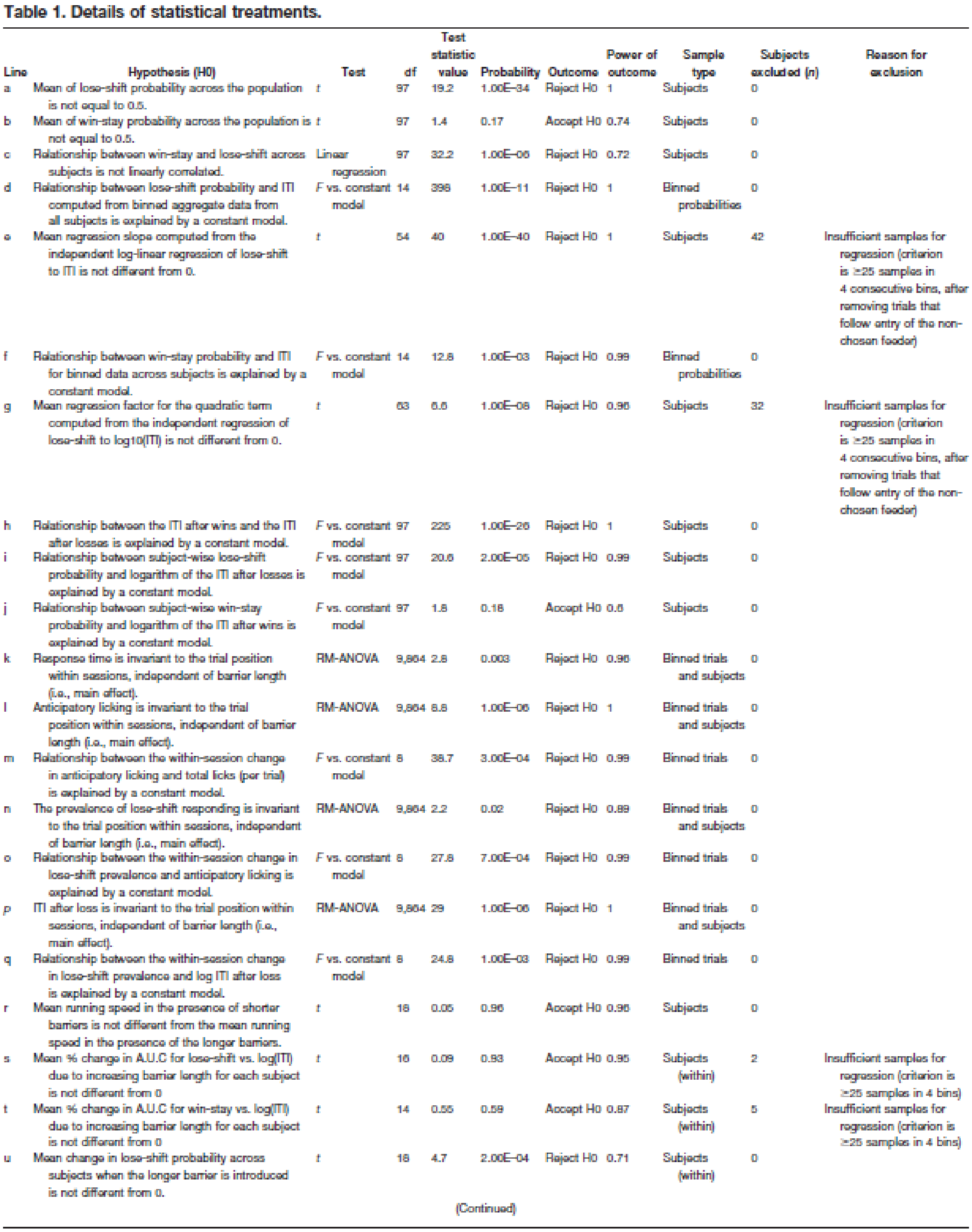

Table 1. Details of statistical treatments.

Line	Hypothesis (H0)	Test	df	Test statistic value	Probability	Outcome	Power of outcome	Sample type	Subjects excluded (*n*)	Reason for exclusion
a	Mean of lose-shift probability across the population is not equal to 0.5.	*t*	97	19.2	1.00E–34	Reject H0	1	Subjects	0	
b	Mean of win-stay probability across the population is not equal to 0.5.	*t*	97	1.4	0.17	Accept H0	0.74	Subjects	0	
c	Relationship between win-stay and lose-shift across subjects is not linearly correlated.	Linear regression	97	32.2	1.00E–06	Reject H0	0.72	Subjects	0	
d	Relationship between lose-shift probability and ITI computed from binned aggregate data from all subjects is explained by a constant model.	*F* vs. constant model	14	398	1.00E–11	Reject H0	1	Binned probabilities	0	
e	Mean regression slope computed from the independent log-linear regression of lose-shift to ITI is not different from 0.	*t*	54	40	1.00E–40	Reject H0	1	Subjects	42	Insufficient samples for regression (criterion is ≥25 samples in 4 consecutive bins, after removing trials that follow entry of the non-chosen feeder)
f	Relationship between win-stay probability and ITI for binned data across subjects is explained by a constant model.	*F* vs. constant model	14	12.8	1.00E–03	Reject H0	0.99	Binned probabilities	0	
g	Mean regression factor for the quadratic term computed from the independent regression of lose-shift to log10(ITI) is not different from 0.	*t*	63	6.6	1.00E–08	Reject H0	0.96	Subjects	32	Insufficient samples for regression (criterion is ≥25 samples in 4 consecutive bins, after removing trials that follow entry of the non-chosen feeder)
h	Relationship between the ITI after wins and the ITI after losses is explained by a constant model.	*F* vs. constant model	97	225	1.00E–26	Reject H0	1	Subjects	0	
i	Relationship between subject-wise lose-shift probability and logarithm of the ITI after losses is explained by a constant model.	*F* vs. constant model	97	20.6	2.00E–05	Reject H0	0.99	Subjects	0	
j	Relationship between subject-wise win-stay probability and logarithm of the ITI after wins is explained by a constant model.	*F* vs. constant model	97	1.8	0.18	Accept H0	0.6	Subjects	0	
k	Response time is invariant to the trial position within sessions, independent of barrier length (i.e., main effect).	RM-ANOVA	9,864	2.8	0.003	Reject H0	0.96	Binned trials and subjects	0	
l	Anticipatory licking is invariant to the trial position within sessions, independent of barrier length (i.e., main effect).	RM-ANOVA	9,864	6.8	1.00E–06	Reject H0	1	Binned trials and subjects	0	
m	Relationship between the within-session change in anticipatory licking and total licks (per trial) is explained by a constant model.	*F* vs. constant model	8	38.7	3.00E–04	Reject H0	0.99	Binned trials	0	
n	The prevalence of lose-shift responding is invariant to the trial position within sessions, independent of barrier length (i.e., main effect).	RM-ANOVA	9,864	2.2	0.02	Reject H0	0.89	Binned trials and subjects	0	
o	Relationship between the within-session change in lose-shift prevalence and anticipatory licking is explained by a constant model.	*F* vs. constant model	8	27.8	7.00E–04	Reject H0	0.99	Binned trials	0	
p	ITI after loss is invariant to the trial position within sessions, independent of barrier length (i.e., main effect).	RM-ANOVA	9,864	29	1.00E–06	Reject H0	1	Binned trials and subjects	0	
q	Relationship between the within-session change in lose-shift prevalence and log ITI after loss is explained by a constant model.	*F* vs. constant model	8	24.8	1.00E–03	Reject H0	0.99	Binned trials	0	
r	Mean running speed in the presence of shorter barriers is not different from the mean running speed in the presence of the longer barriers.	*t*	18	0.05	0.96	Accept H0	0.96	Subjects	0	
s	Mean % change in A.U.C for lose-shift vs. log(ITI) due to increasing barrier length for each subject is not different from 0	*t*	16	0.09	0.93	Accept H0	0.95	Subjects (within)	2	Insufficient samples for regression (criterion is ≥25 samples in 4 bins)
t	Mean % change in A.U.C for win-stay vs. log(ITI) due to increasing barrier length for each subject is not different from 0	*t*	14	0.55	0.59	Accept H0	0.87	Subjects (within)	5	Insufficient samples for regression (criterion is ≥25 samples in 4 bins)
u	Mean change in lose-shift probability across subjects when the longer barrier is introduced is not different from 0.	*t*	18	4.7	2.00E–04	Reject H0	0.71	Subjects (within)	0	

(Continued)

Table 1. Continued

Line	Hypothesis (H0)	Test	df	Test statistic value	Probability	Outcome	Power of outcome	Sample type	Subjects excluded (*n*)	Reason for exclusion
v	Mean difference between predicted and actual lose-shift decrease due to increased barrier length is not different from 0.	*t*	18	0.14	0.89	Accept H0	0.95	Subjects (within)	0	
w	Mean change in rewarded trials due to barrier length is not different from 0.	*t*	18	2.45	0.02	Reject H0	0.92	Subjects (within)	0	
x	The prevalence of lose-shift responding is invariant to the trial position within sessions, independent of barrier length (i.e., main effect).	RM-ANOVA	6,109	1.6	0.16	Accept H0	0.42	Binned trials and subjects	0	
y	The ITI after loss is invariant to the trial position within sessions, independent of barrier length (i.e., main effect).	RM-ANOVA	6,109	5.7	3.00E−05	Reject H0	0.99	Binned trials and subjects	0	
z	Anticipatory licking is invariant to the trial position within sessions, independent of barrier length (i.e., main effect).	RM-ANOVA	6,109	6.8	4.00E−06	Reject H0	1	Binned trials and subjects	0	
aa	The prevalence of lose-shift responding is invariant to barrier length, independent of changes due to trial position in the session (i.e., main effect).	RM-ANOVA	1,18	8.3	0.01	Reject H0	0.78	Binned trials and subjects	0	
ab	The ITI after loss is invariant to barrier length, independent of changes due to trial position in the session (i.e., main effect).	RM-ANOVA	1,18	28	5.00E−05	Reject H0	1	Binned trials and subjects	0	
ac	Anticipatory licking is invariant to barrier length, independent of changes due to trial position in the session (i.e., main effect).	RM-ANOVA	1,18	0.5	0.52	Accept H0	0.9	Binned trials and subjects	0	
ad	Relationship between lose-shift responding and anticipatory licking is explained by a constant model.	*F* vs. constant model	5	10.1	0.02	Reject H0	0.58	Binned trials	0	
ae	Mean difference in win-stay probability across subjects computed after a previous win vs. two previous wins at the same feeder is not greater than 0.	*t*	48	10.2	1.00E−13	Reject H0	1	Subjects (within)	2	Insufficient occurrence of win-stay-wins sequences (criterion is ≥25)
af	Mean difference in lose-shift probability across subjects computed after a previous loss vs. two previous losses at the same feeder is not greater than 0.	*t*	32	2.2	0.99	Accept H0	1	Subjects (within)	18	Insufficient occurrence of lose-stay-lose sequences (criterion is ≥25)
ag	Mean prediction accuracy of the Q-learning model and win-stay-lose-shift is not different from 0.	*t*	34	5.2	1.00E−05	Reject H0	0.96	Subjects	0	
ah	The median probability of lose-shift on the second training session is not different from chance (0.5).	Wilcox	17		0.03	Reject H0	0.77	Subjects	0	
ai	Mean probability of lose-shift did not change across training or testing days.	RM-ANOVA	15,150	0.54	0.91	Accept H0	1	Subjects, sessions	0	
aj	Mean probability of win-stay did not change across training or testing days.	Wilcox	17		0.01	Reject H0	0.83	Subjects	0	
ak	Mean probability of win-stay did not change across training or testing days.	RM-ANOVA	15,150	2.3	5.00E−03	Reject H0	1	Subjects, sessions	0	

Table 2. Main behavioural results (mean ± 95% confidence interval). Asterisks (*) indicate *p* < 0.05.

Experiment	Group	N	Win-stay	Lose-shift	Response time
Lateral habenula lesions	Controls	13	0.65 ± 0.05	0.67 ± 0.05	1.53 ± 0.22
	LHb lesions (> 45% lesioned)	8	0.49 ± 0.09*	0.68 ± 0.05	1.13 ± 0.11*
Systemic amphetamine injections (Repeated measures)	Saline injections	7	0.71 ± 0.04	0.81 ± 0.04	1.71 ± 0.17
	Systemic amphetamine injections (1.5 mg/kg)	7	0.61 ± 0.07*	0.78 + 0.06	1.63 ± 0.26*
Nucleus accumbens core amphetamine microinfusions (Repeated measures)	Saline microinfusions	8	0.61 ± 0.06	0.70 ± 0.05	1.94 ± 0.37
	Amphetamine microinfusions (40 µg/µl)	8	0.52 ± 0.11*	0.80 ± 0.05	1.56 ± 0.44*

FIGURES

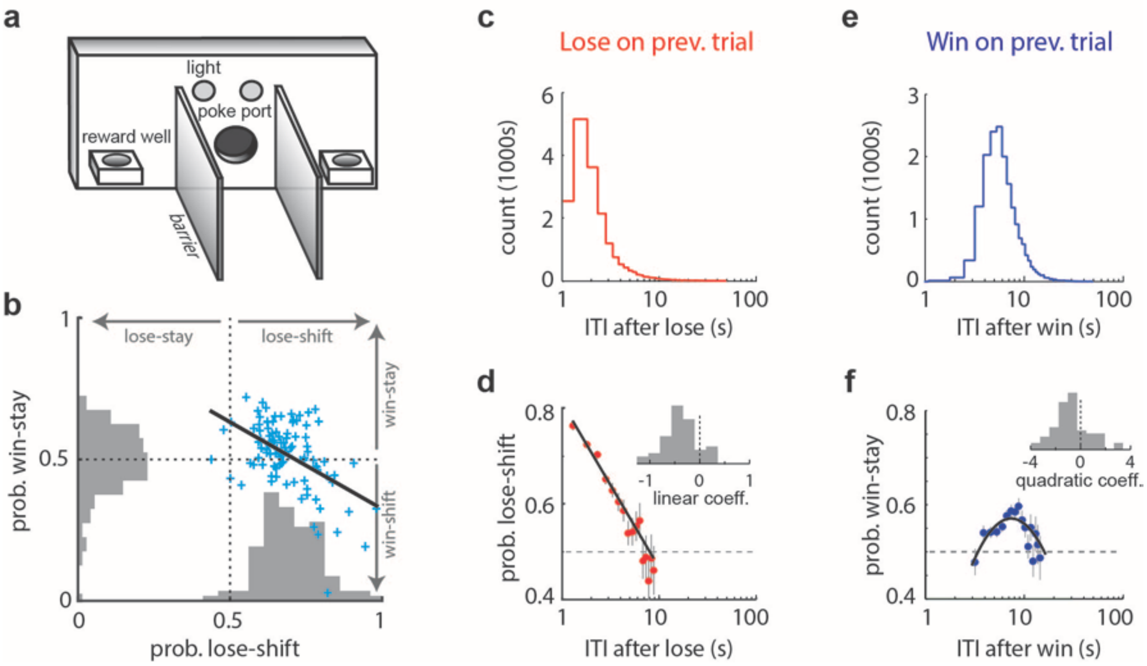

Figure 1. Prevalence of win-stay and lose-shift responses. A, Schematic illustration of the behavioural apparatus. B, Scatter plot and population histograms of win-stay and lose-shift responding, showing that these strategies are anticorrelated among subjects. C, Frequency of ITIs after loss trials across the population. D, Probability of lose-shift computed across the population for the bins of ITI in C, revealing a marked log-linear relationship. Individual subjects also exhibit this behaviour, as indicated by the nonzero mean of the frequency histogram of linear coefficient terms for fits to each subject’s responses (inset; see text for statistical treatment). E, F, Plots for win-stay analogous to those in C and D reveal a log-parabolic relationship with ITIs in the population and individual subjects. Vertical lines in D and F indicate SEM, and the dashed lines indicate chance levels (Prob = 0.5).

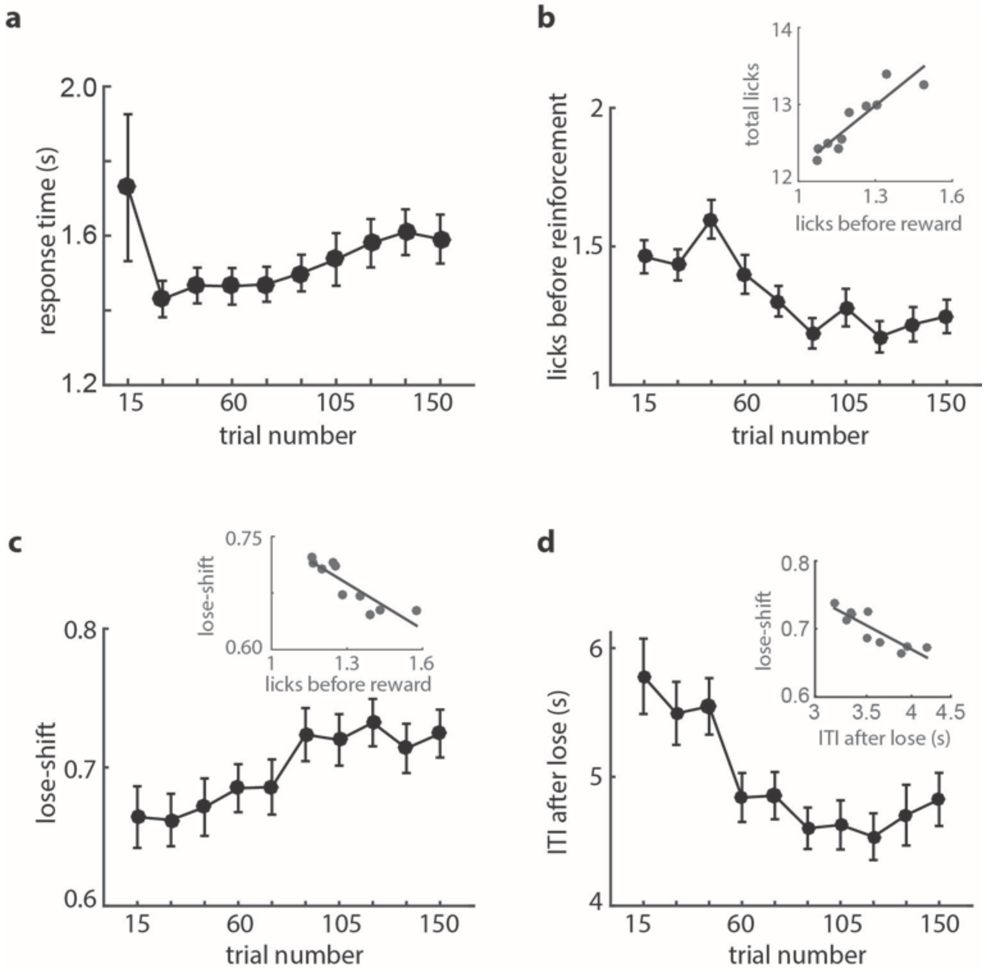

Figure 2. Within-session changes of dependent variables. A, Mean response time (from nose-poke to feeder) over 15 consecutive trials and all animals in Figure 1. Response time increases throughout the session after trial 30, suggesting a progressive decrease in motivation. B, Mean number of licks before reinforcement, which decreases within the session. The number of these anticipatory licks correlates strongly with the total number of licks at each feeder within the session (inset). C, Mean probability of lose-shift, which increases within the session and negatively correlates with licking (inset). D, Mean ITI after loss trials decreases within session. The within-session variance of lose shift correlates strongly with the log of the within-session ITI after losses (inset). Error bars indicate SEM.

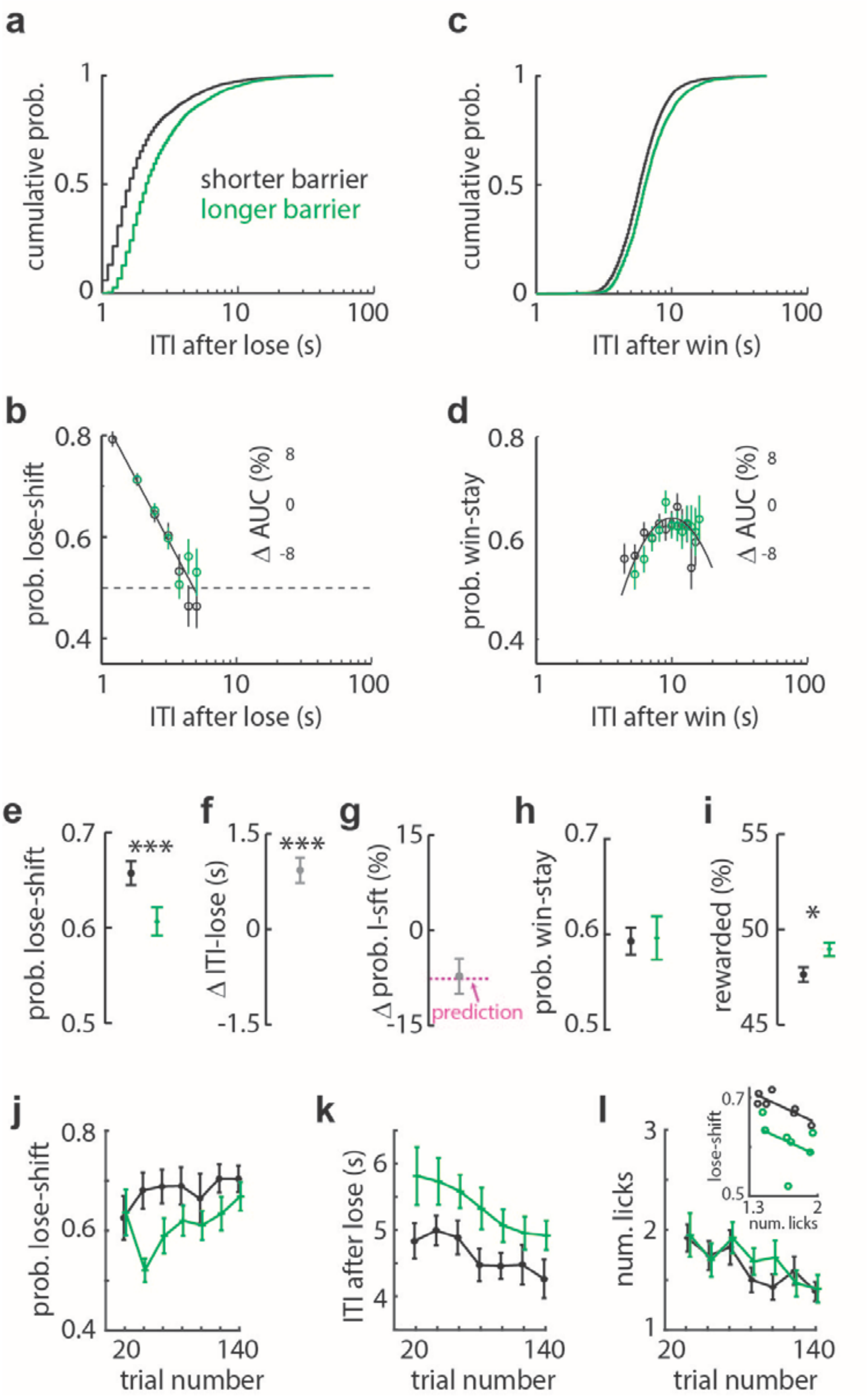

Figure 3. Invariance of lose-shift and win-stay models to movement times. A, Frequency of population ITIs after losses showing that intervals were increased

for long (green) compared with short (dark) barriers. B, Probability of lose-shift computed across the population independently for short (dark) and long (green) barriers. Both conditions were fitted well by the common model (dark solid line). The change in the area under the curve computed independently for each subject between conditions shows no difference (inset), indicating that the mnemonic process underlying lose-shift responding is invariant to the ITI distribution. C, D, Plots of ITI and probability of stay responses after wins, showing that win-stay is also invariant to barrier length. E, Mean lose-shift responding across subjects is decreased by longer barriers. F, Within-subject ITI increases after loss trials under long barriers compared with short barriers. G, Mean within-subject change in the probability of lose-shift due to longer barriers is predicted (magenta dashed line) by the change in ITI based on the log-linear model. H, Mean probability of win-stay computed across animals is not altered by barrier length. I, Long barriers led to more rewarded trials per session because of the reduction in predictable lose-shift responding. J, Mean probability of lose-shift for bins of 20 trials and rats for long and short barriers, showing an increase across sessions for either barrier length. K, Mean ITI after loss for each barrier condition, showing a decrease within the session. L, Mean number of licks prior to reinforcement across the session, showing a decrease within sessions but no effect of barrier length. (L, inset) Plots of lose-shift and licking for each barrier condition, showing that licking is not sufficient to account for variance in lose-shift between barrier conditions. Statistically significant difference among group means: *$p < 0.05$, *** $p < 0.001$. Error bars show SEM.

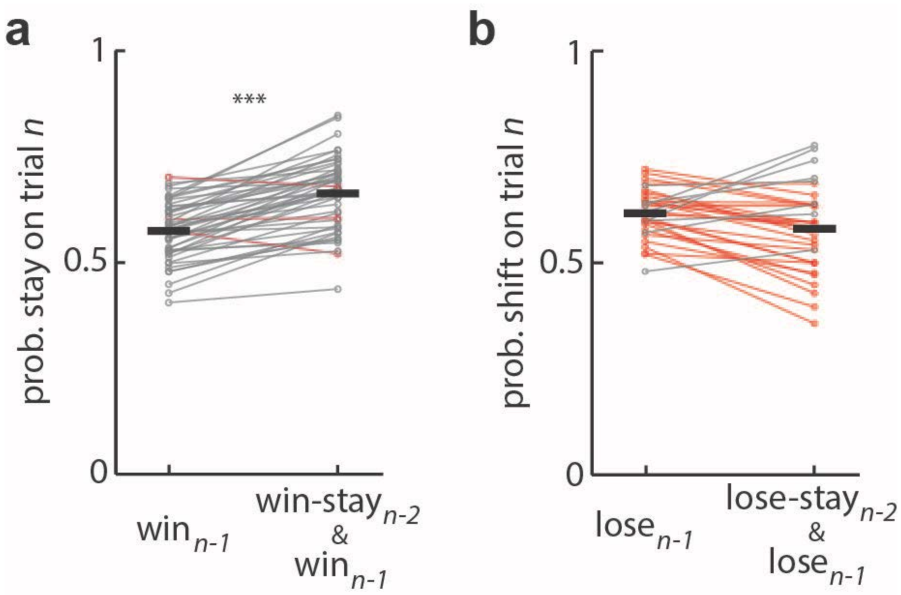

Figure 4. Effect of consecutive wins or losses on choice: test for reinforcement learning. A, Plot of probability of a stay response on trial n, after a win (i.e., win-stay; left) or win-stay-win sequence (right) for each rat. The latter is the probability that the rat will chose the same feeder in three consecutive trials given wins on the first two of the set. The data show an increased probability of repeating the choice given two previous wins on the same feeder compared with a win on the previous trial, consistent with RL theory. B, Plot of probability of a switch response on trial n after a loss (lose-shift; left) or after a lose-stay-lose sequence (right). The probability of shifting after two consecutive losses to the same feeder is not greater than the probability of shifting after a loss on the previous trial, which is inconsistent with the predictions of RL theory. In both plots, gray lines indicate a within-subject increase in probability, whereas red lines indicate a decrease. ***Statistical significance of increased probability ($p < 0.001$) within subjects.

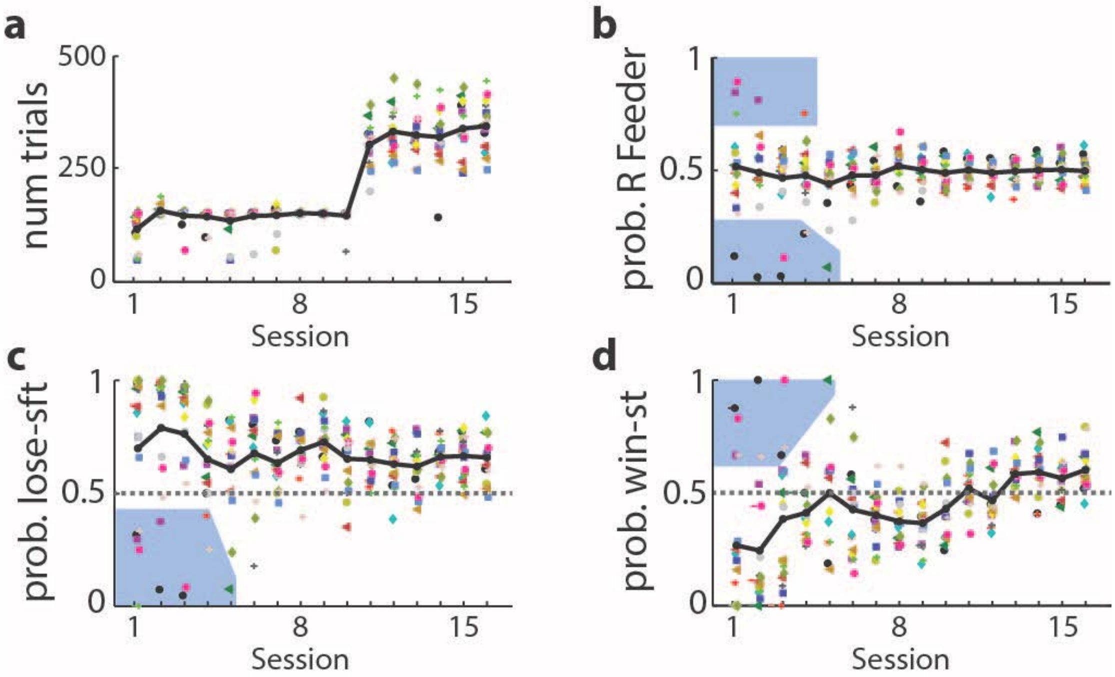

Figure 5. Responses during every training session for one cohort. Responses plotted for each rat (symbol-color) and each day of training. Session 1 is the second time the rats were placed into the behavioural box, and reward probability was p = 0.5 for each feeder regardless of previous responses or rewards. A, Number of trials completed in each session. Rats were allowed 90 min to complete up to 150 trials in sessions 1–10, and hallways of increasing lengths were introduced in sessions 3–8. B–D, Plot of the probability of responding to the rightward feeder, probability of lose-shift, and probability of win-stay during the first 16 sessions. The majority of rats showed no side bias, strong lose-shift, and very little win-stay in initial trials. Only a few rats showed initial side bias, and therefore little lose-shift and strong win-stay (blue shading in panels B–D). Lose-shift was invariant over training, whereas win-stay increased (see text). Dark lines indicate median across all subjects for each day.

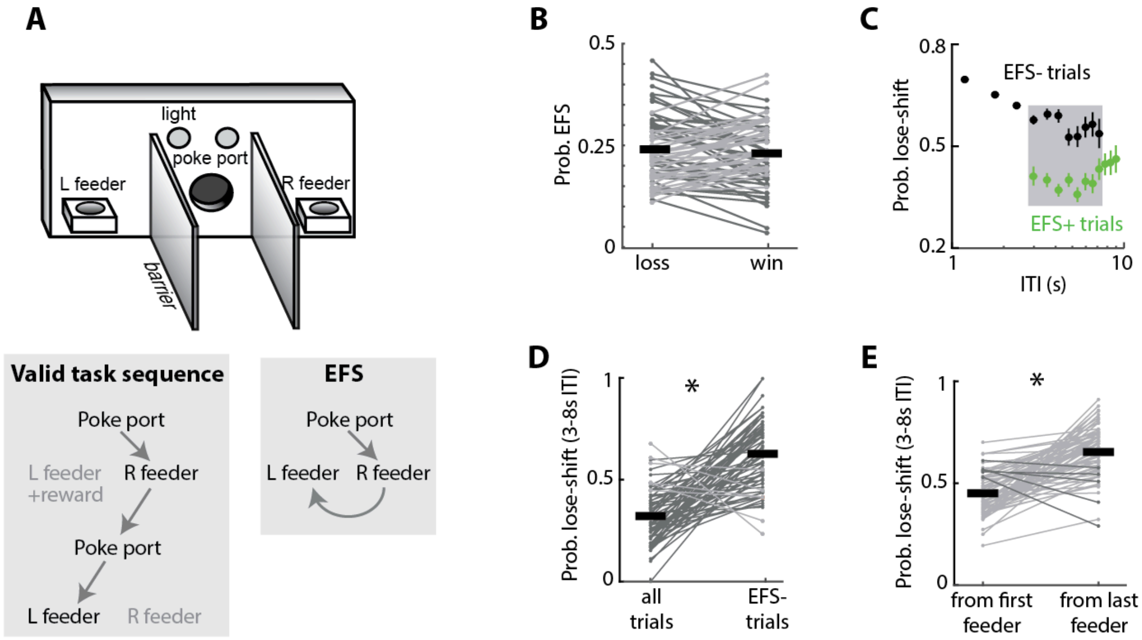

Figure 6. Task apparatus and responding. (A) Schematic representation of the behavioural apparatus and examples of operant sequences on the task. Valid sequences consist of a nose poke in the poke port followed by locomotion to one of the two feeders. Rats sometimes chose to locomote from one feeder to the other without committing a nose poke; we term this extraneous feeder sampling (EFS). (B) The probability of EFS immediately after reward (win) or reward omission (loss) for each rat (Cohort 1: n = 68 for this and subsequent panels), showing that reinforcement does not affect EFS likelihood. (C) The probability of lose-shift responding following trials with EFS (green) or no EFS (black) parsed into bins of inter-trial-interval. EFS dramatically reduces lose-shift probability regardless of ITI for the population. (D) The within-subject plot of mean lose-shift probability. (E) Mean lose-shift probability for each rat computed from either the first feeder chosen after the nose poke or the last feeder chosen before the subsequent nose poke. Nearly all rats appeared to generate lose-shift responses from the last feeder chosen as compared to the first feeder chosen, suggesting that the EFS strongly influences subsequent choice. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from the comparison group ($p < 0.000001$).

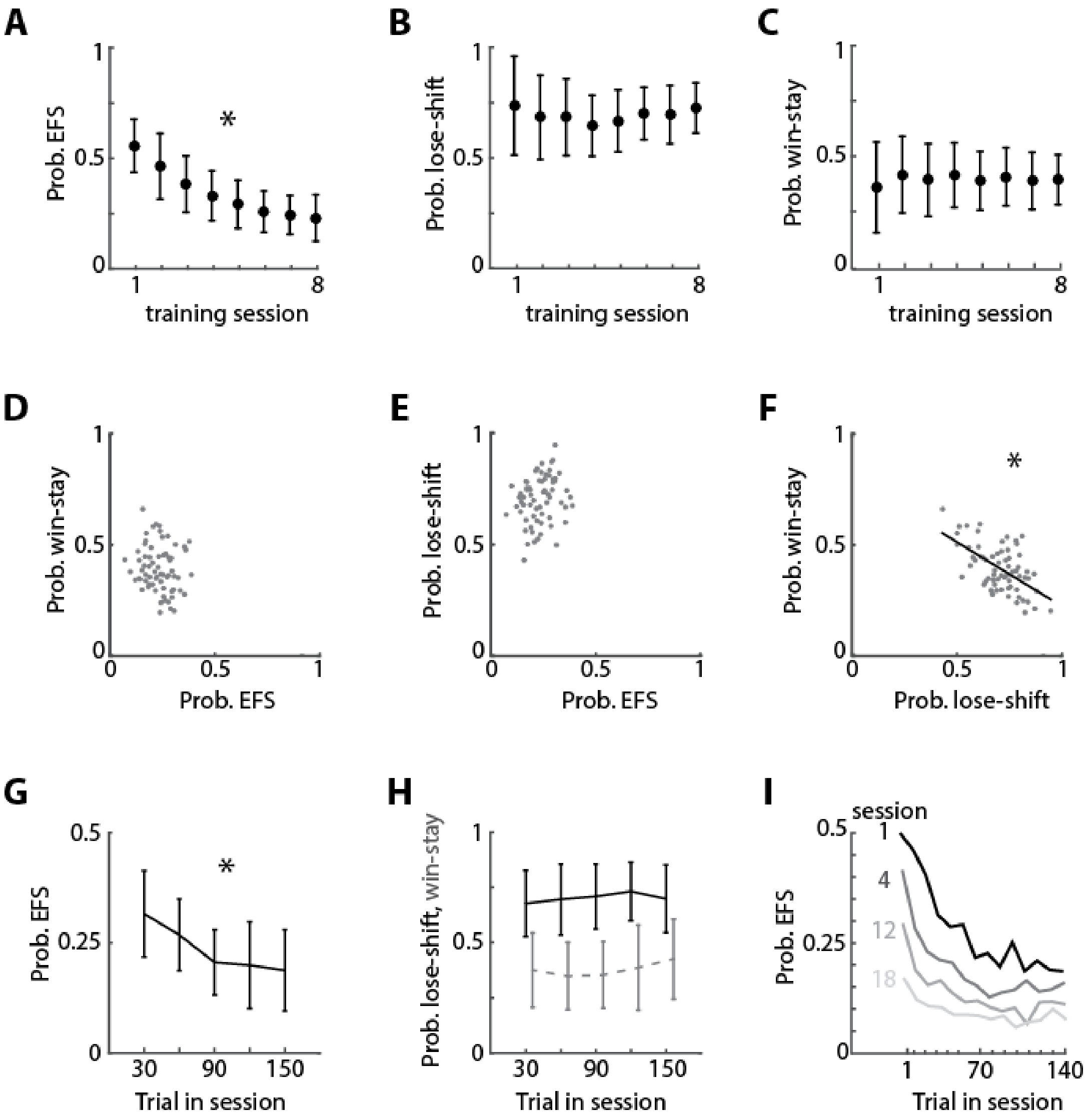

Figure 7. Changes in EFS and win-stay-lose-shift responding within and between sessions. (A) The mean probability of EFS decreases over the training sessions (Cohort 1: n = 68). (B, C) The mean probability of lose-shift or win-stay responding does not change over the training sessions. (D-F) Correlations among the probability of EFS, lose-shift, and win-stay responding among rats on the last day of training. The immediate effect of EFS on win-stay and lose-shift measures were minimized by omitting trials following EFS. EFS was uncorrelated with the other response types. (G-H) The plot of the probability of EFS, lose-shift, and win-stay (dashed line) for bins of 30 trials within sessions. Only EFS decreased within sessions. (I) The plot of EFS probability versus trial bin (10 trials/bin) within each of several sessions of a separate cohort of rats (Cohort 2, n = 30), showing that within-session variance of EFS reduces with training. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from the comparison group ($p < 0.000001$).

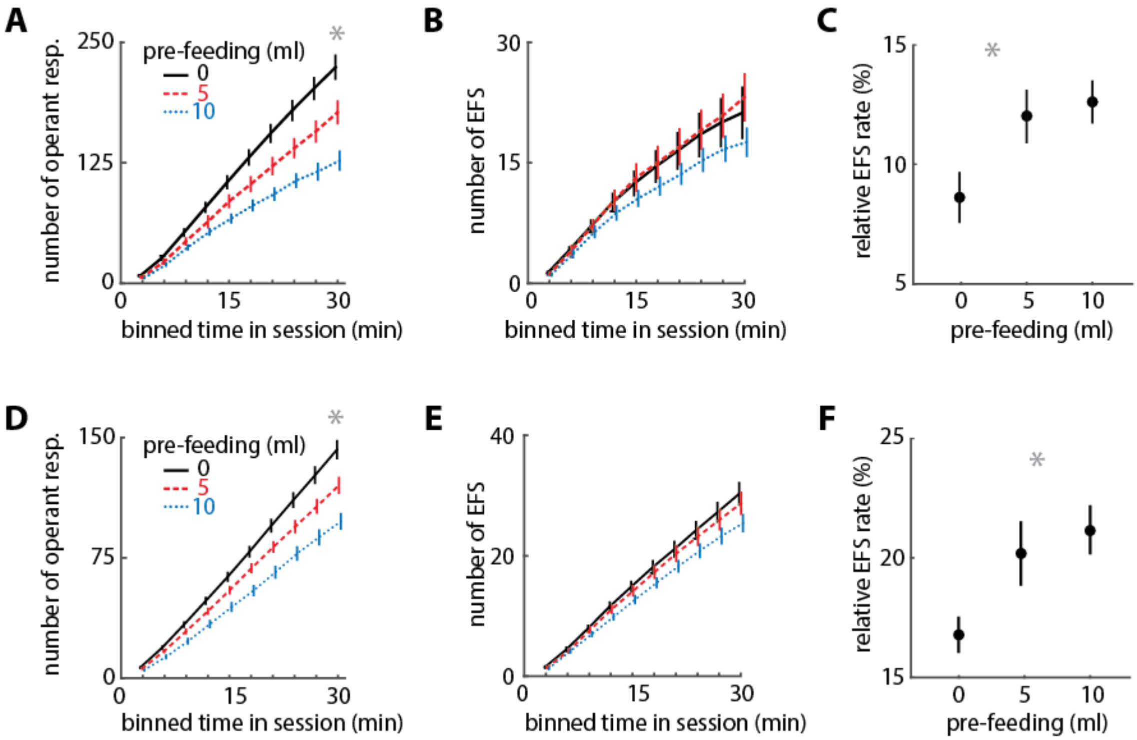

Figure 8. Effect of devaluation on task performance. (A) Mean cumulative sum of operant responses (nose-poke to feeder) in bins of time within a session (Cohort 2: n = 30 rats in panels A-C). Pre-feeding rats 20 minutes before the task reduced the number of trials performed. (B) The mean cumulative sum of EFS events in the same sessions, which was not reduced by pre-feeding. (C) The mean relative rate of EFS/trials for each pre-feeding level, showing an increase with devaluation. (D-F) Same plots as above for a new heterogeneous cohort collected by different experimenters (Cohort 3: n = 48 in panels D-F), showing replication of the devaluation effects. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from the comparison group ($p < 0.003$).

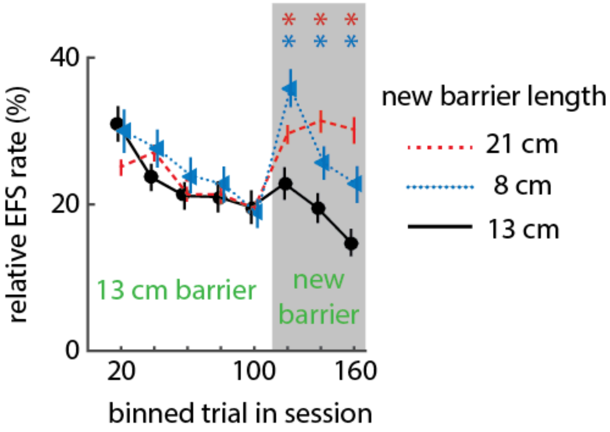

Figure 9. Effect of mid-session change in the barrier on EFS rate. Mean relative EFS rate within a session before and after the barrier was replaced at trial 101 (male subjects from Cohort 3: n =16). Replacing the barrier with either a longer (red dashed line) or shorter (blue dotted line) barrier increased EFS as compared to replacing with the same length barrier (black solid line). Asterisks (*) indicate a significant difference of means by post-hoc analysis (P < 0.04).

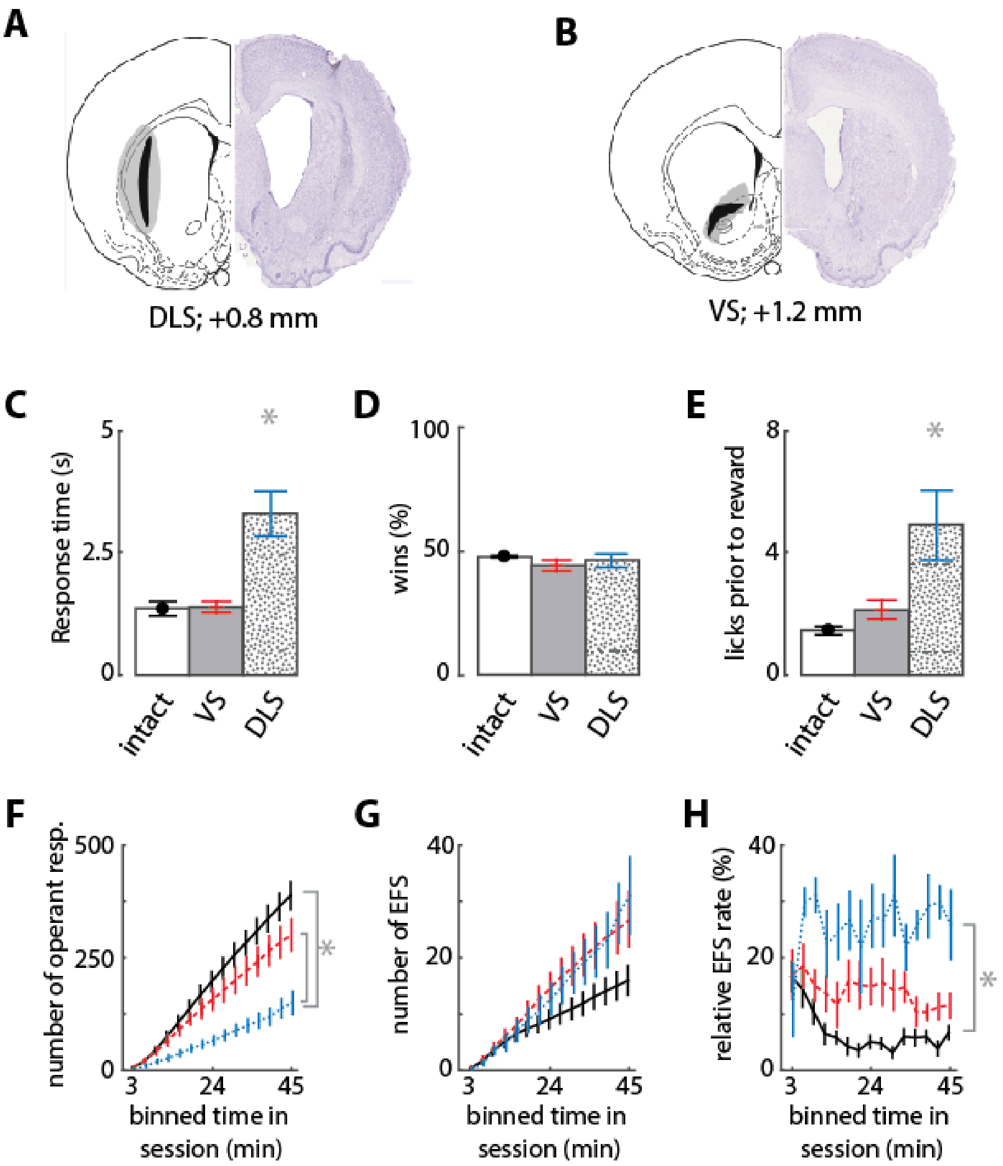

Figure 10. Effects of lesions of the dorsolateral striatum (DLS) or nucleus accumbens core (NACc). (A-B) The extent of the excitotoxic striatal lesions in Cohort 4 (n = 21). The black and grey shading show minimal and maximal extent of the lesions to the dorsolateral striatum (DLS, n = 7) or nucleus accumbens core (NACc, n = 7), respectively. (C) Mean response times, showing that DLS-lesioned rats made slower responses than the other two groups (n = 7 controls). (D) The mean percentage of rewarded trials was not affected by lesions. (E) Mean number of licks prior to reward. (F) Cumulative sum of completed task trials in bins of time within sessions. (G) Cumulative sum of EFS, showing no reduction in lesioned rats relative to controls. (H) Relative rate of EFS to operant responses (trials) within sessions, showing a dramatic increase in DLS-lesioned animals. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from each other by Tukey's HSD post-hoc test (p < 0.01).

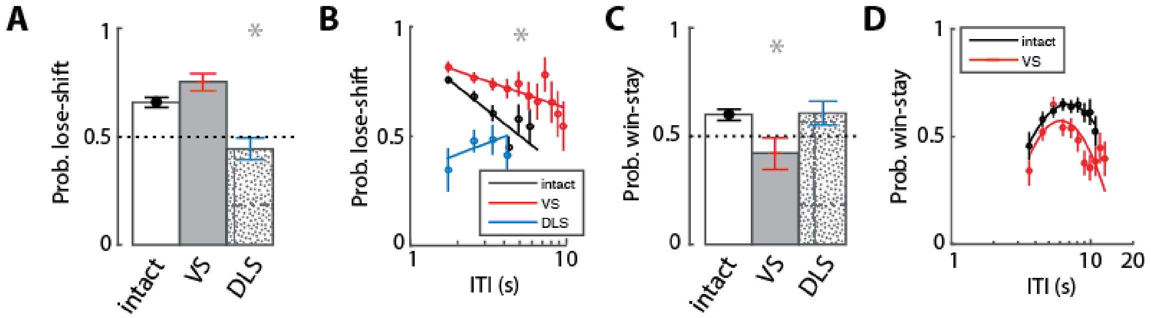

Figure 11. Effect of striatal lesions on lose-shift and win-stay responding. (A) The group-averaged probability of lose-shift responding, showing that DLS-lesioned animals decreased lose-shift relative to the other groups, and approached the optimal level in this task (p = 0.5). (B) The plot of the probability of lose-shift versus the logarithm of the inter-trial-interval for each group. The DLS-lesioned animals show low lose-shift regardless of the ITI. (C) The group-averaged probability of win-stay responding. (D) The plot of the probability of win-stay versus the logarithm of the inter-trial-interval, showing that animals with NACc lesions have reduced win-stay probabilities regardless of ITI. Error bars indicate SEM, and asterisks (*) indicate group means that were significantly different from each other by Tukey's HSD post-hoc test (p < 0.05).

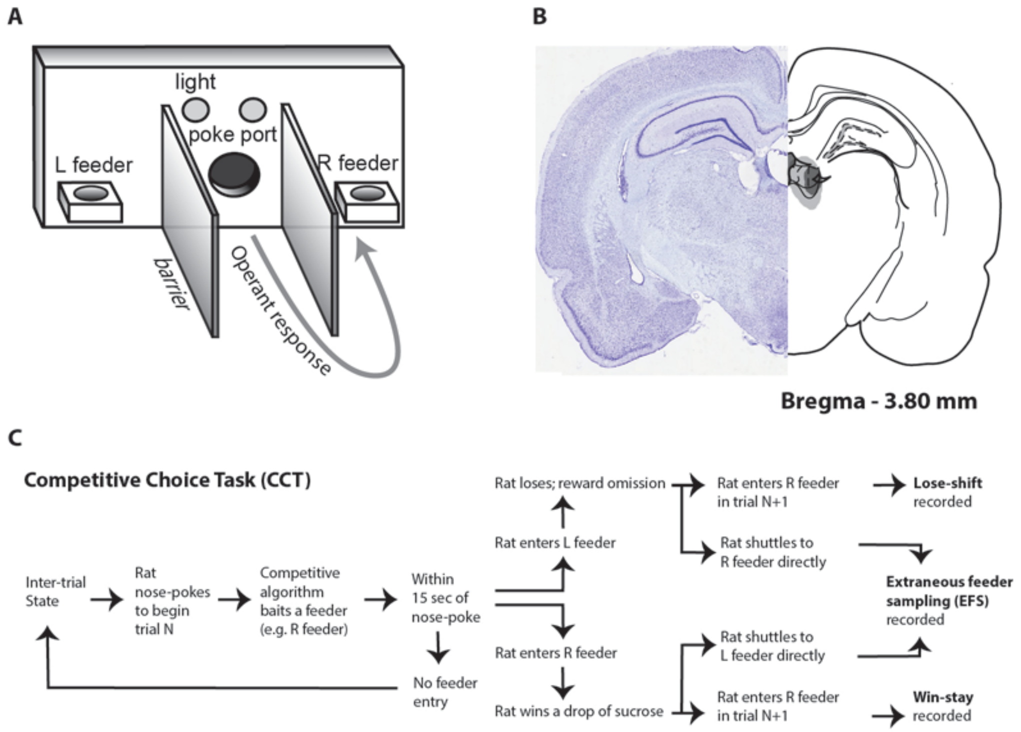
A
light
poke port
L feeder
R feeder
barrier
Operant response
B
Bregma - 3.80 mm
C
Competitive Choice Task (CCT)
Inter-trial State
Rat nose-pokes to begin trial N
Competitive algorithm baits a feeder (e.g. R feeder)
Within 15 sec of nose-poke
No feeder entry
Rat enters L feeder
Rat loses; reward omission
Rat enters R feeder in trial N+1
Lose-shift recorded
Rat shuttles to R feeder directly
Extraneous feeder sampling (EFS) recorded
Rat enters R feeder
Rat wins a drop of sucrose
Rat shuttles to L feeder directly
Rat enters R feeder in trial N+1
Win-stay recorded

Figure 12. Behavioural task and histology. (A) Schematic representation of the competitive choice task apparatus and the operant response. (B) A representative Nissl-stained brain section along with a schematic representation of the largest (light grey) and smallest (dark shading) electrolytic lesions of the lateral habenula. (C) flowchart of the sequence of events and the definition of lose-shift, win-stay, and extraneous feeder approach (EFS) in the competitive choice task.

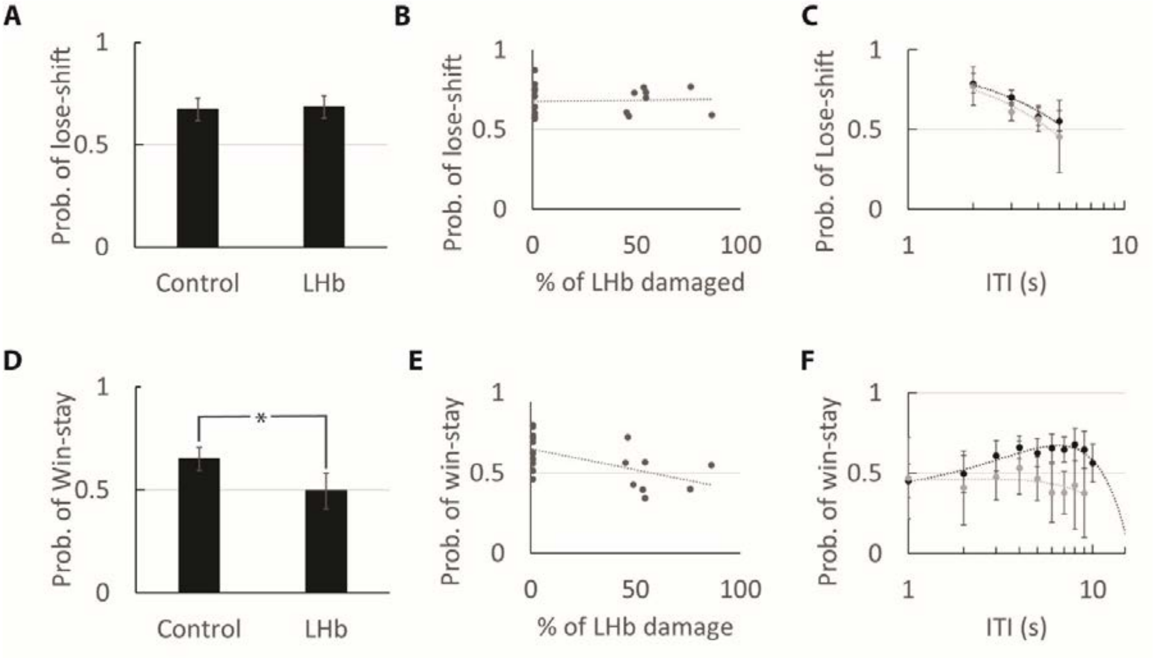

Figure 13. Effects of LHb lesions on win-stay and lose-shift responses in the competitive choice task. (A) Probability of lose-shift responding by the two experimental groups on the test day. (B) Correlation between the probability of lose-shift responding and the stereological estimates of the percentage of LHb damaged in each subject. Animals plotted at 0% are controls. r = 0.056. (C) Probability of lose-shift responding as a function of inter-trial-interval (ITI). The black dotted lines represent data from the control group, whereas the grey dotted lines represent the LHb lesion group. (D) Probability of win-stay responding. (E) Correlation between the probability of win-stay responding and the stereological estimates of the percentage of LHb damaged in each subject. r = -0.582. (F) Probability of lose-shift as a function of ITI. Error bars represent 95% confidence intervals, and asterisks (*) indicate statistically different means ($p < 0.05$).

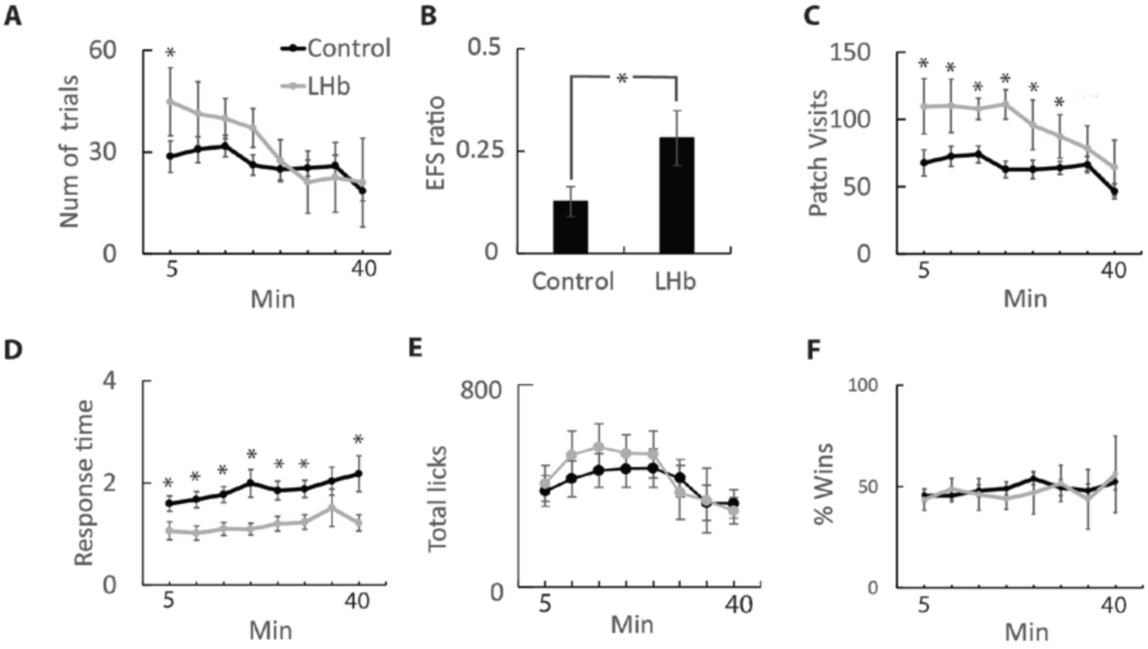

Figure 14. Effects of LHb lesions on locomotion in the competitive choice task. (A) Number of trials completed within a session in 5-min time bins. (B) Ratio of extraneous feeder sampling (EFS) to the number of operant trials. (C) Number of 'patch visits', which is the sum of the number of distinct entries made to the feeders and nose-port. (D) Average time taken to reach a feeder after making a nose-poke. (E) Total number of licks made at either of the feeder wells. (F) Percentage of trials that were rewarded with sucrose. Error bars represent 95% confidence intervals, and asterisks (*) indicate statistically different means ($p < 0.05$).

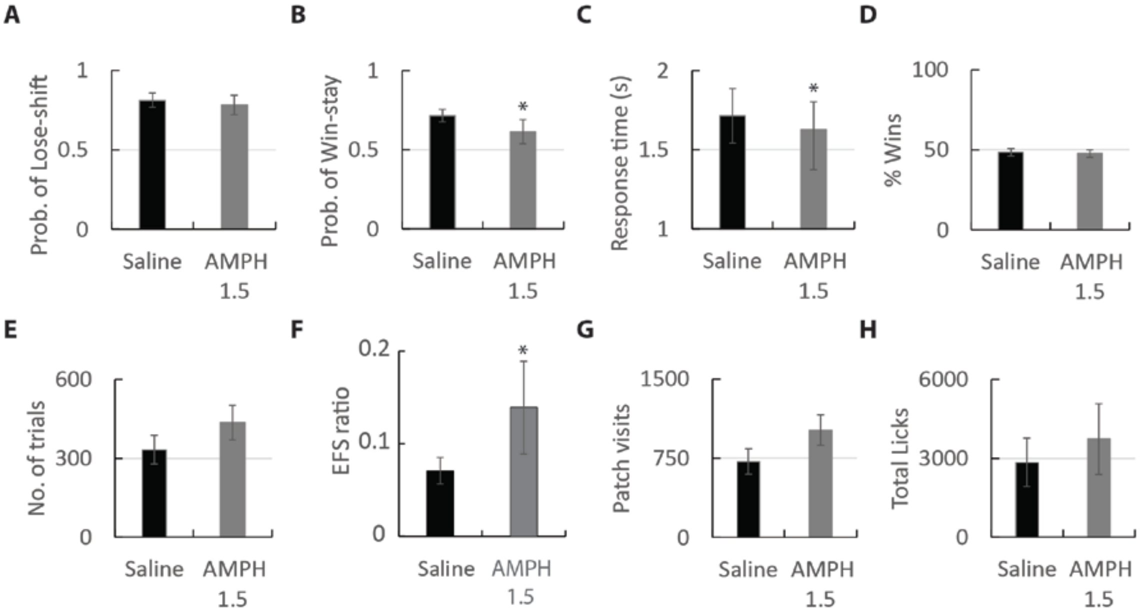

Figure 15. Effects of acute systemic d-amphetamine administration. (A) Probability of lose-shift responding. (B) Probability of win-stay responding, which is decreased following AMPH injections. (C) Average response time taken by the rats to locomote from the nose-poke port to one of the feeder wells, showing that rats are faster on AMPH. (D) Percentage of trials in which the animals were rewarded. (E) Number of operant trials completed within the test session. (F) Ratio of extraneous feeder sampling (EFS) to the number of operant trials. (G) Number of patch visits. (H) Total number of licks made in the feeder wells within a 45-min session. ‘AMPH 1.5’ indicates the amphetamine dosage of 1.5 mg/kg. Error bars represent 95% confidence intervals, and asterisks (*) indicates significantly different means ($p < 0.05$) computed by repeated measures ANOVA.

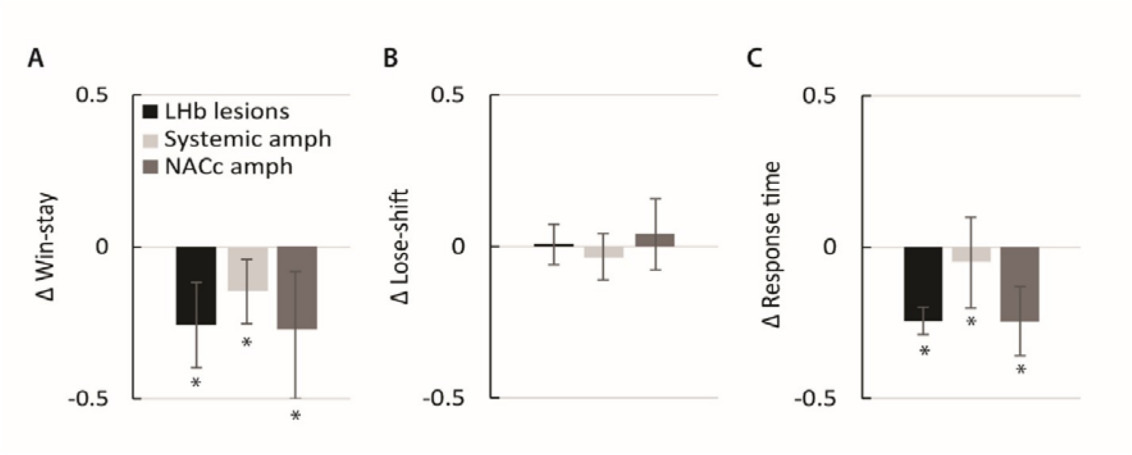

Figure 16. Comparison of behavioural changes resulting from lateral habenula lesion (LHb lesions), systemic AMPH, and AMPH microinfusion into the nucleus accumbens core (NACc AMPH; previously reported). The bars represent the difference between the experimental and control group in each respective study. (A) Average change in the probability of win-stay response. (B) Average change in the probability of lose-shift response. (C) Average change in response time. Error bars represent 95% confidence intervals, and asterisks (*) indicate significant differences of means in the experimental group compared to their respective controls ($p < 0.05$) determined by ANOVA.

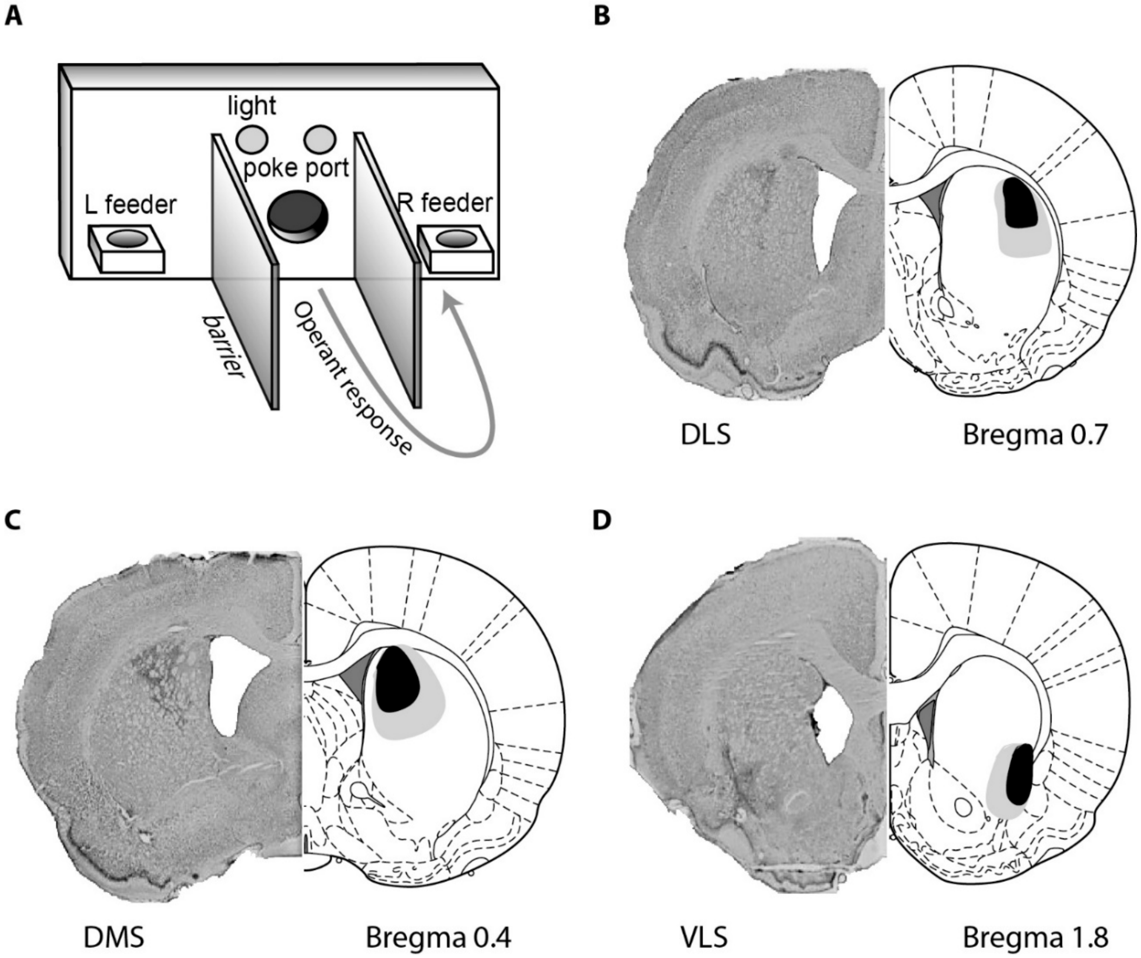

Figure 17. Behavioural task and histology. (A) Schematic representation of the competitive choice task apparatus and the operant response. (B-D) Representative Nissl-stained brain sections along with a schematic representation of the largest (light grey) and smallest (dark shading) excitotoxic lesions of the dorsolateral striatum (DLS), dorsomedial striatum (DMS), and ventrolateral striatum (VLS) in rats.

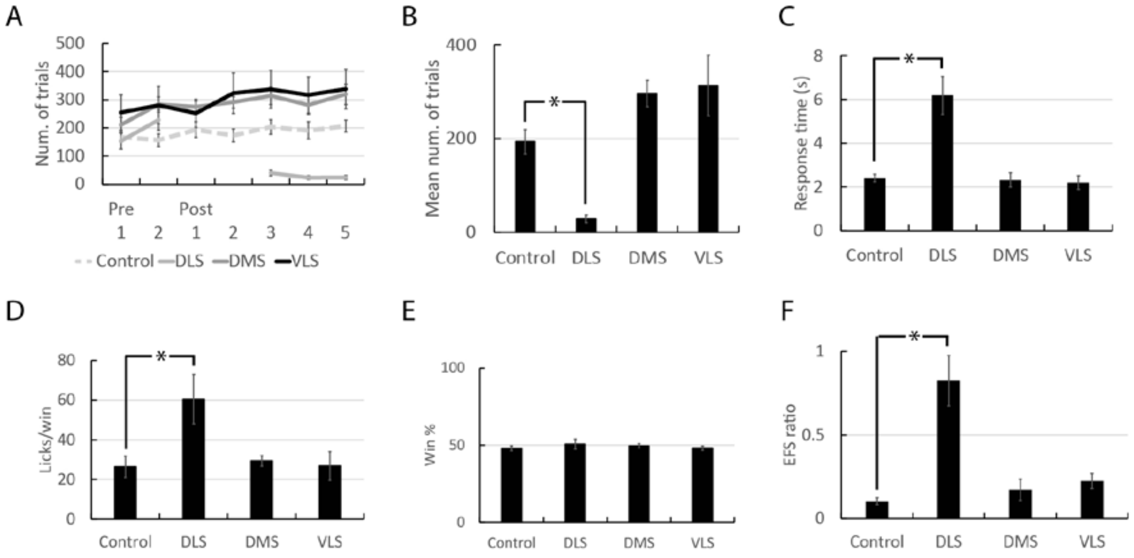

Figure 18. Effects of the striatal subregion lesions on motor function and motivation. (A) Mean number of trials completed in each session including the two sessions prior to surgery (Pre) and the five testing sessions after lesions (Post). (B) Mean number of trials completed in the post-surgery sessions. (C) Average time taken to reach a feeder after making a nose-poke. (D) The average ratio of total licks over total wins in a session. (E) Percentage of trials that were rewarded with sucrose. (F) The ratio of EFS to the number of operant trials. Error bars represent SEM and asterisks (*) indicate statistically different means based on Tukey post-hoc test (p < 0.05).

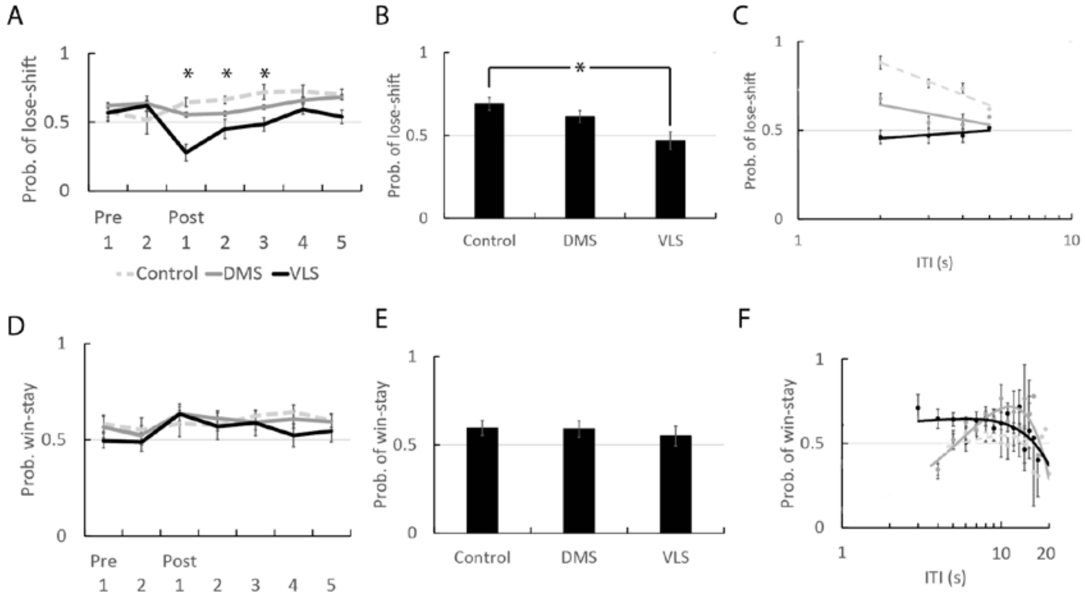

Figure 19. Effects of striatal lesions on win-stay and lose-shift behaviour. (A) The probability of lose-shift responding in each session including two last sessions before surgery (Pre) and five testing sessions after lesions (Post). (B) Average probability of lose-shift responding in the post-surgery sessions. (C) The probability of lose-shift responding as a function of inter-trial-interval (ITI). (D) The probability of win-stay responding in each session. (E) Average probability of win-stay responding in the post-surgery sessions. (F) The probability of win-stay as a function of ITI. Error bars represent SEM and asterisks (*) indicate statistically different means based on Tukey post-hoc test ($p < 0.05$).