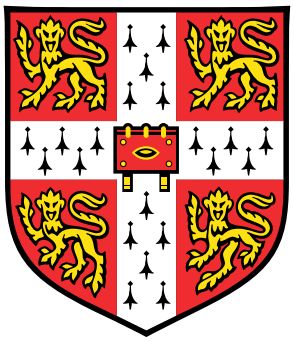
Neuromodulation and neural circuits underlying cognitive flexibility in the rat



Júlia Sala-Bayo

Department of Psychology
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Jesus College

September 2020

Declaration

The work described in this dissertation was carried out between October 2016 and September 2020 at the Department of Psychology, University of Cambridge under the supervision of Professor Jeffrey W. Dalley and Professor Trevor W. Robbins. Data for Chapter 5 and 6 were collected at the CNS Department at Boehringer Ingelheim, Germany, between October 2018 and September 2019.

This dissertation has resulted from my own work and all collaborations are specified in the text. This dissertation is not substantially the same as any that I have submitted, or is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University of similar institution, except as specified in the text.

I have made every attempt to reference properly for any idea or finding that is not my own. The length of this dissertation does not exceed the word limit of the School of Biology degree committee i.e. 60,000 words.

Júlia Sala-Bayo
September 2020

Abstract

Cognitive flexibility refers to how individuals adapt their behaviour to changes in the environment. Although important for survival and wellbeing, cognitive flexibility is impaired in a wide range of neurological and neuropsychiatric disorders, including Parkinson's Disease and Obsessive-Compulsive Disorder (OCD). Optimal flexibility is known to depend on dopamine (DA) neurotransmission in the central nervous system, but the precise mechanism and brain loci underlying the effects of DA on flexible decision-making remain unclear.

In this thesis, cognitive flexibility was inferred in experimental rats by evaluating their performance on a reversal-learning task involving a simple discrimination between rewarded and non-rewarded stimuli. During reversal, subjects must adapt and respond to the formerly non-rewarded stimulus whilst ignoring the initially rewarded stimulus. Learning on this task thus requires constants shifts in behaviour in response to positive (rewarded) and negative (non-rewarded) feedback. The overarching hypothesis of my thesis is that DA modulates reversal learning performance by signaling positive and negative reward prediction errors (RPE) within the direct (rewarded) and indirect (non-rewarded) pathways, respectively.

To investigate this hypothesis, I used a range of experimental approaches to interrogate the neuromodulation of the direct and indirect pathways by DA. In Chapter 3, I found dissociable effects of D1 and D2 receptor antagonists during different stages of serial visual reversal learning when administered into the nucleus accumbens shell. In Chapter 4, I used a recently developed valence-probe visual discrimination task to dissociate different components of reversal learning and tested the extent to which these were dependent on D2 receptors. We found that the D2 agonist quinpirole impaired reversal learning when given systemically, an effect that depended on decreased sensitivity to negative feedback, but improved performance when given directly into the nucleus accumbens. In Chapter 5, the synaptic location of D2 receptors involved in the modulation of reversal learning was evaluated. Using a post-synaptic probe compound (an adenosine 2A receptor antagonist) evidence is presented for a

predominately post-synaptic locus underlying the effects of D2 agents on reversal learning. Finally, in Chapter 6, an *in-vivo* optogenetics intervention was used to simulate activity in the mesoaccumbal and nigrostriatal circuits during reversal learning. Here, activation of the mesoaccumbal, not nigrostriatal, circuit modulated reversal learning on trials when the expected reward was omitted.

Taken together, these original results provide support for a dissociable role of DA receptors and striatal subregions in learning from positive and negative feedback in reversal learning. These findings expand our understanding of the neural circuit mechanisms underlying cognitive inflexibility and highlight potential therapeutic targets to improve flexible decision making in PD, OCD and a range of other brain disorders.

Named publications relating to this PhD thesis

- **Sala-Bayo, J.**, Fiddian, L., Nilsson, S. R. O., Hervig, M. E., McKenzie, C., Mareschi, A., Boulos, M., Zhukovsky, P., Nicholson, J., Dalley, J. W., Alsiö, J.*, Robbins, T. W.* (2020). Dorsal and ventral striatal dopamine D1 and D2 receptors differentially modulate distinct phases of serial visual reversal learning. *Neuropsychopharmacology*, 45:736-744. (**Chapter 3**).

- Alsiö, J., Phillips, B. U., **Sala-Bayo, J.**, Nilsson, S. R. O., Calafat-Pla, T. C., Rizwand, A., Plumbridge, J. M., Lóopez-Cruz, L., Dalley, J. W., Robbins, T. W. (2019). Dopamine D2-like receptor stimulation blocks negative feedback in visual and spatial reversal learning in the rat: behavioural and computational evidence. *Psychopharmacology*, 236:2307-2323. (**Chapter 4**).

- **Sala-Bayo, J.**, Alsiö, J., Selin, M., Baghurst, E. B., Wilson, M. E., Nicholson, J., Robbins, T. W., Dalley, J. W. Pharmacological evidence that dopaminergic D2 receptors in the nucleus accumbens modulate improvements in reversal learning performance in rats. *In preparation*. (**Chapter 4**).

- **Sala-Bayo, J.**, Piller, S., Deiana, S., von Heimendahl, M., Nicholson, J., Robbins, T. W., Alsiö, J.*, Dalley, J. W.*. Adenosine 2A receptors blockade reverses the impairment in reversal learning induced by D2 receptors agonism and antagonism: a behavioral and computational approach. *In preparation*. (**Chapter 5**).

- **Sala-Bayo, J.**, Piller, S., Nissen, W., von Heimendahl, M., Deiana, S., Nicholson, J., Robbins, T. W., Alsiö, J.*, Dalley, J. W.*. Optogenetic stimulation of mesoaccumbal, but not nigrostriatal, projections blocks learning from losses in reversal learning. *In preparation*. (**Chapter 6**).

* These authors contributed equally to this work.

Acknowledgements

First, I would like to thank my supervisors, Jeff Dalley, Trevor Robbins, and Janet Nicholson, for giving me this opportunity, and for all your guidance and support throughout these years. It has been a pleasure and privilege working with and learning from you.

I would also like to thank my advisor, Johan Alsiö, for your daily advice and inspiration. Your passion for your research was the first thing to hook me into doing this PhD.

I would like to thank everybody in the Dalley-Robbins lab and second floor, past and present, who made my time in the department so enjoyable. Chiara Toschi, I am glad I got share to the PhD adventure with you from the very first day. Thank you for your questioning spirit and cheerful attitude. Discussing our present and future steps with you has been priceless. Laura López-Cruz, thank you for your invaluable friendship and professional advice. Louise Piilgaard, thank you for being an amazing source of motivation and inspiration. You are a star. Mathilda Selin, you made my arrival to Cambridge so sweet, and made working late hours much more fun. People like you are only found *once in a blue moon*. Parisa Moazen, thank you for being so caring and embodying the true meaning of teamwork. Peter Zhukovsky, I really appreciate your genuine interest for my results, and for your wonderful help with computational modelling. Ben Phillips, thanks for your scientific advice and chats in long journeys to conferences. Katharina Zuhlsdorf, thank you for your positive attitude and for always being keen to help. Thanks also to Mona Hervig, Karly Turner, Jolyon Jones, Colin McKenzie, Chiara Giuliano, Mickael Puaud, and all of the CAF staff for your support; you have all contributed to making this adventure much better.

I have been fortunate to learn from and collaborate with a number of scientists in Boehringer Ingelheim, Germany, including Serena Deiana, Moritz van Heimendahl, and Wiebke Nissen. Thank you for all your guidance, supervision, and learning experience. I would also like to acknowledge the excellent assistance I received from Andrea Blasius, Nathalie Okogun, Anna Kaun, and Sammy Piller that helped in making this project possible.

I would also like to thank the girls in JCBC W1, and the crew from ‘Tipsy Tuesdays’. You kept me sane during these years, and made this experience amazing. A special thank you to my ‘Hermanas Salesianas’, for your friendship and support throughout these years, for making me feel like there was always a bit of home with me wherever I was, and for really meaning it when you said that ‘distance means so little when someone means so much’.

I would like to thank my family for believing in me, and understanding that the distance is sometimes necessary.

Finally, the biggest thank you to Alex for everything over the past years. I am very grateful I have been so lucky to spend my time in Cambridge with you.

Table of contents

Abbreviations

List of Abbreviations and Acronyms

5-HT 5-hydroxytryptamine; Serotonin

6-OHDA 6-hydroxydopamine

A- Negative stimulus A (unrewarded)

A Action

A+ Positive stimulus A (rewarded)

A2AR Adenosine 2A receptors

aDLS Anterior dorsolateral striatum

aDMS Anterior dorsomedial striatum

AL After a loss

A-O Action-outcome

AP Anteroposterior

ASL After a spurious win

ASW After a spurious loss

ATP Adenosine triphosphate

AW After a win

AWERB Animal Welfare and Ethical Review Body

B- Negative stimulus B (unrewarded)

B+ Positive stimulus B (rewarded)

BLA Basolateral amygdala

$C_{50/50}$ Stimulus rewarded 50% of the times

C Context

cAMP Cyclic-adenosine monophosphate

CANTAB Cambridge Automated Neuropsychological Test Automated Battery

CG Cingulate gyrus

ChR2 Channelrhodopsin 2

CN Central nucleus of the amygdala

CNS Central nervous system

cP Centipoise

CR Conditioned response

CS Conditioned stimulus

CS- Negative conditioned stimulus (unrewarded)

CS+ Positive conditioned stimulus (rewarded)

D1R Dopamine D1-like receptors

D2L Dopamine D2 receptor long isoform

D2R Dopamine D2-like receptors

D2S Dopamine D2 receptor short isoform

D3R Dopamine D3 receptors

D4R Dopamine D4 receptors

D5R Dopamine D5 receptors

DA Dopamine

DAT Dopamine transporter

DAT Dopamine transporter

dH Dorsal hippocampus

DLS Dorsolateral striatum

DMS Dorsomedial striatum

DREADDs Designed receptors exclusively activated by designed drugs

DSM-5 Diagnostic and Statistical Manual of Mental Disorders

DV Dorsoventral

ENT Entorhinal cortex

GABA γ-aminobutyric acid

GP Globus pallidus

GPe External part of the globus pallidus

GPi Internal part of the globus pallidus

HPB Kleptose hydroxypropyl β-cyclodextrin

i.p. Intraperitoneal

IL Infralimbic cortex

LSQ Latin-Square

ML Mediolateral

mPFC Medial prefrontal cortex

NAc Nucleus accumbens

NAcC Nucleus accumbens core

NAcS Nucleus accumbens shell

NGS Normal goat serum

NIMH National Institute of Mental Health

O Outcome

OCD Obsessive compulsive disorder

OFC Orbitofrontal cortex

p.o. Per os

P Pallidum

PBS Phosphate-buffered saline

pDMS Posterior dorsomedial striatum

PFC Prefrontal cortex

PIT Pavlovian-instrumental transfer

PL Prelimbic cortex

PP Parietal cortex

PRL Probabilistic reversal learning

Q Quinpirole

R Raclopride

RT-PCR Reverse transcriptome polymerase chain reaction

S Stimulus

SMA Sensorimotor cortex

SN Substantia nigra

SNc Substantia nigra pars compacta

SNr Substantia nigra pars reticulata

S-O Stimulus-outcome

S-R Stimulus-response

SSRI Selective serotonin reuptake inhibitor

STN Subthalamic nucleus

TH Tyrosine hydroxylase

UR Unconditioned response

US Unconditioned stimulus

UUC Up until choice

vH Ventral hippocampus

VMAT2 Vesicular monoamine transporter 2

VPVD Valence-probe visual discrimination

VTA Ventral tegmental area

WCST Wisconsin Card Sorting Test

WGTA Wisconsin General Testing Apparatus

WO Washout

Z ZM-241385

Chapter 1

Introduction

1.1 Cognitive flexibility

Cognitive flexibility refers to the skill to adapt behaviour to sudden changes in the environment. Since cognitive flexibility is assessed by observed behaviour, it is often referred to as behavioural flexibility when assessed in experimental animals. Eslinger and Grattan (1993) described behavioural flexibility as *"the ability to shift avenues of thought and action in order to perceive, process and respond to situations in different ways"*. Behavioural flexibility is an emergent property of efficient executive functioning that serves many independent, but interacting, cognitive control processes including attention, working memory, inhibition, and shifting, which are necessary for adaptive behaviour in novel or unfamiliar environments (Dajani and Uddin, 2015).

Despite flexible behaviour being important for everyday life, this capacity is often restricted in a number of neurological and neuropsychiatric disorders, including schizophrenia (Leeson et al., 2009), obsessive-compulsive disorder (OCD) (Remijnse et al., 2013), Parkinson's disease (Cools et al., 2001), and substance use disorder (Ersche et al., 2011). Such cognitive impairment is often treatment-resistant or even worsens with pharmacotherapy (Pallanti et al., 2004). To find treatments for these patients, it is critical to understand the pathophysiology and neural mechanisms of impaired behavioural flexibility.

The neural substrates of behavioural flexibility are most commonly assessed in humans and other animals using reversal learning tasks. In such tasks, an initially acquired action-

outcome (A-O) association (e.g. responding to stimulus A in an operant chamber) is paired with a rewarded outcome, while responding to stimulus B is paired with lack of reward. When the discrimination is successfully learnt according to a criterion, the reward contingencies are reversed such that stimulus A is no longer rewarded and *vice versa*. Optimal reversal performance requires the A-O association to be flexibly updated to facilitate the density of rewarded outcomes (Izquierdo et al., 2017; Izquierdo and Jentsch, 2012; Nilsson et al., 2015). A difficulty in disengaging from a previously rewarded stimulus that is now unrewarded reflects rigid behaviour (Fineberg et al., 2010; Izquierdo and Jentsch, 2012; Nilsson et al., 2015). The associative structure of reversal learning is deceptively complicated with fluctuations in behaviour dependent on A-O learning, together with stimulus-outcome (S-O) and stimulus-response (S-R) learning. In addition, several cognitive processes are recruited to enable adaptive learning in new conditions, comprising associative learning, response selection and inhibition, decision-making, working memory, and attention.

The brain substrates that underlie behavioural flexibility have been vigorously pursued in recent years (Izquierdo, 2017). Accumulating evidence from rodents, non-human primates and humans has linked inflexible behaviour in reversal learning to dysfunction within the cortico-striatal circuitry, involving connections to the dorsal and ventral striatum in rodents (Boulougouris et al., 2007; Dalton et al., 2016; McAlonan and Brown, 2003), with widely researched modulatory contributions from dopamine (DA) and serotonin (5-HT) (Clarke et al., 2004, 2011).

1.1.1 The relevance of cognitive flexibility research for neuropsychiatric disorders

Neuropsychiatric and neurological disorders have long been studied but we still lack a comprehensive understanding of their underlying neural mechanisms, as well as broad and efficient treatments. Symptom heterogeneity, complexity and co-morbidities commonly observed in mental disorders, represent a challenge in psychiatry to appropriately diagnose and treat patients. Neurocognitive endophenotypes (i.e. measurable biomarkers correlated with a cognitive illness), have therefore received much attention within the last decade (Fineberg et al., 2010; Flint and Munafò, 2007; Robbins et al., 2012) in which behavioural inflexibility has been proposed as an endophenotype for a host of heterogeneous neuropsychiatric disor-

ders including OCD, schizophrenia, and Parkinson's disease (Gillan et al., 2016; Robbins et al., 2012).

Splitting heterogeneous disorders into cognitive endophenotypes can promote translational research across species, enabling us to better dissociate commonalities between disorders and optimise diagnosis. This could explain apparent comorbidities across disorders, and more importantly, lead to tailor-made pharmacological and behavioural treatments with a transdiagnostic application (Robbins et al., 2012). Indeed, the National Institute of Mental Health (NIMH) has recently endorsed the new strategy of focusing on behavioural constructs, rather than symptoms and disorder classifications stated by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (Association, 2013) in clinical and preclinical research. The existence of an inflexible endophenotype shared by a range of heterogeneous cognitive disorders suggests a common neural mechanism that could potentially be targeted for drug remediation (Fineberg et al., 2010; Godier and Park, 2014; Izquierdo et al., 2017; Romera-Garcia et al., 2020).

For instance, OCD is a heterogeneous disorder characterised by maladaptive patterns of repetitive and inflexible behaviour and cognition. Deficits in cognitive flexibility have been reported in individuals with OCD and their unaffected first-degree relatives during reversal learning (Chamberlain et al., 2007; Gruner and Pittenger, 2017; Gu et al., 2008). During reversal cognitive tests, the orbitofrontal cortex (OFC) and the striatum show attenuated responsiveness, suggesting they are key brain regions in modulating the symptoms underlying the behavioural impairment and being key in the neurobiological mechanism of OCD (Remijnse et al., 2006). Schizophrenia is a disorder characterised by accentuated positive or psychotic symptoms alongside deficits in emotion, motivation and cognition (Waltz and Gold, 2016). Specially the latter has been associated with low goal-directed performance due to slow acquisition of adaptive behaviour and flexible responses following a change in contingencies (Morris et al., 2015; Waltz and Gold, 2016). Consistent with this clinical observation, decreased cognitive flexibility is found in reversal learning tasks after impairments in the cortico-striatal circuit (Leeson et al., 2009; Morris et al., 2015; Reddy et al., 2016). Individuals with schizophrenia also perform poorly on outcome devaluation tasks commonly used to assess habitual behaviour (Morris et al., 2015). Parkinson's disease affects the initiation and control of movements, motivation and reward-seeking behaviour (Borek et al., 2006). A classic neuropathology of this disease is the degeneration of DA cells in the substancia nigra pars compacta (SNc) impairing both tonic and phasic DA signalling in its efferent targets (Dauer and Przedborski, 2003). In consequence, individuals with Parkinson's

disease experience poor levels of performance in reversal learning (Peterson et al., 2009), which can be altered depending on DA medication (Cools et al., 2001).

Understanding the neural mechanisms underlying these behaviours and subprocesses is crucial to shed light into the aetiology of neuropsychiatric disorders and contribute to finding effective treatments.

1.2 Reversal learning

Reversal learning is a widely used procedure to assess cognitive flexibility and has been broadly used to investigate aberrant cognitive processing associated with neuropsychiatric disorders (Fig. 1.1). Reversal learning measures the ability to adapt behaviour to a reversal in reinforcement contingencies. Initially, subjects are trained to discriminate between two (or more) stimuli or locations, one of which is associated with reward, whereas the other one is not. After successfully discriminating both options by reaching a criterion level of performance, contingencies reverse, so that the previously rewarded stimulus is now non-rewarded, and *vice versa*. Subjects are then trained to reach the criterion again. This next phase requires acquiring a strategy to solve the task during initial discrimination, which must be inhibited or extinguished when reversal occurs to prevent perseverative (i.e. incorrect) responses. Once the old strategy is extinguished, subjects need to acquire a novel association to be rewarded in the new conditions. That is, reversal learning requires a shift in valence between stimuli or locations that have previously been associated with a specific outcome (e.g. delivery of reward). Importantly, reversal learning can consist in a reversal of all type of cues (e.g. visual, spatial, odorant, textural), but the choosing options remain constant.

1.2.1 Reversal learning

Reversal learning paradigms can be used in multiple species, including rodents (Graybeal et al., 2011), non-human primates (Clarke et al., 2007), and humans (Cools et al., 2007), and therefore have inherent translational utility to resolve the neural and psychological mechanisms of cognitive flexibility.

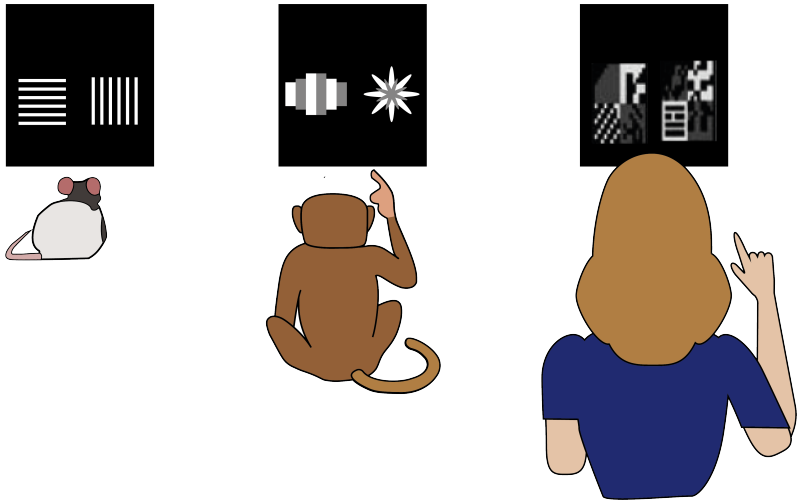


Fig. 1.1 Illustration of the reversal-learning task using touchscreens in rodents, monkeys and humans. Adapted from Izquierdo et al. (2017).

In rodents, reversal-learning procedures often performed in operant chambers comprising levers, apertures for nose poke responses, or a touch-sensitive screen – although mazes have also been adopted (Bari et al., 2010; Dalton et al., 2014; Shah et al., 2019). Reversal learning can use purely spatial stimuli, like lever or response apertures, or incorporate visual and auditory cues (Boulougouris et al., 2007; Castañé et al., 2010). Touchscreens are mainly used to test visual-discrimination reversal learning since they offer a wider variety of stimuli, but they can also test spatial strategies (Alsiö et al., 2019; Mar et al., 2013; Oomen et al., 2013). Touchscreen testing is analogous to the procedures used in the Cambridge Automated Neuropsychological Test Automated Battery (CANTAB) (Robbins et al., 1994; Sahakian et al., 1993), a set of computerised neuropsychological tests developed to assess cognitive flexibility in humans.

In non-human primates, the tasks used are similar to those used in rodents, but in general monkeys are able to solve more reversals than rodents within the same time period (Dalton et al., 2014; Horst et al., 2019). In addition, modified versions of the Wisconsin General Test Apparatus (WGTA) can be implemented in monkeys to investigate reversal learning. In the WGTA, two wells are presented and covered by an object. One of the two wells contains a reward, which the monkeys discover by removing the correct object. Similarly, it can be tested by using visual stimuli or cards (Jones and Mishkin, 1972; McAlonan and Brown, 2003; Walker et al., 2009).

In humans, including healthy volunteers and patients, reversal-learning tests follow the same basic design as the tests mentioned above; i.e. the simultaneous presentation of two stimuli, one of which is associated with reward, while the other one is not. Stimuli are normally presented on a touchscreen or form letters or symbols on a keyboard, where subjects must register their choices. Although tasks are similar across species, the main difference in humans relies on the type of reward, which is often feedback on the correctness of their choices or through receipt of real or hypothetical monetary rewards, instead of food incentives as in the case of experimental animals. Thus, in humans, stimuli are generally conditioned reinforcers rather than primary reinforcers.

Reversal learning paradigms are usually either (1) deterministic ($P(reward|choice) = 1$ or $P(reward|choice) = 0$), or (2) probabilistic ($0.5 < P(reward|choice) < 1$ or $0 < P(reward|choice) < 0.5$) in nature. Probabilistic strategies slow down learning and reduce the development of simple strategies, such as win-stay and lose-shift, since subjects must integrate the history of choices and outcomes to choose which stimulus is more likely to deliver reward. Depending on the particular task configuration, it is also possible for a single reversal over multiple sessions or multiple reversals within the same session (Alsiö et al., 2019).

Reversal learning performance is typically assessed by the number of reversals achieved or by the number of errors *versus* correct responses made before reaching the discrimination criterion. The tendency of subjects to perseverate on the previously rewarded stimulus (now non-rewarded) is a further variable of interest, that reflects a failure to disengage from the previous strategy. Trial-by-trial responses to positive and negative feedback can also be assessed by calculating win-stay and lose-shift probabilities (Alsiö et al., 2019; Bari et al., 2010; Dalton et al., 2014).

More recently, trial-by-trial computational models have attempted to simulate how subjects learn about environmental contingencies and translate reward representations into action (Daw, 2009; Niv et al., 2012). These models have highlighted latent variables such as learning rate, the tendency to explore or exploit stimuli according to learned reinforcing properties, and 'stickiness' – the likelihood to respond on the same stimulus as in previous trials regardless of its rewarding value.

1.2.2 Other paradigms to test cognitive flexibility

In addition to reversal learning, other paradigms have been developed to assess cognitive flexibility, including attentional set shifting, task switching and the ability to suppress automatically elicited responses.

Attentional set shifting falls in the same category as reversal learning in terms of changes in reward contingencies: once subjects have learned an initial discrimination, contingencies shift, so that what was previously positive is now negative and *vice versa*. However, the main difference in attentional set shifting and strategy shifting is that the shift occurs between different dimensions or perceptual categories e.g. from visual to spatial. These tasks allow for intradimensional and extradimensional shifts. In intradimensional tests, the set of stimuli changes, but not the relevant stimulus dimension (e.g. if the first set was based on discrimination visual stimuli, the shift will incorporate novel choice options within the visual domain. In extradimensional tests, not only the set of stimuli shifts, but also the reinforced dimension (e.g. if the first set of stimuli were visual, now they might be auditory).

As in reversal learning, attentional set-shifting paradigms have been developed across species (Eagle et al., 2008; Roberts et al., 2007). In humans, one of the most prominent tasks to assess the ability to shift is the Wisconsin Card Sorting Test (WCST). The WCST consists of presenting to the subject a number of stimulus cards with certain shapes and colours. Subjects are asked to match these cards according to a specific stimulus aspect. Attentional set-shifting tasks for animals have been developed based on the WCST and using sets of stimuli belonging to multiple sensory dimensions (e.g. odorant, visual, spatial, auditory) (Birrell and Brown, 2000; Garner et al., 2006).

Another paradigm used widely to evaluate cognitive flexibility is task switching. It consists of changing the stimulus-response set following an external cue. This is a fundamental difference from reversal learning and attentional set shifting, in which changes are not cued, hence subjects need to explore the conditions to detect the change and develop a new strategy to solve the task. This paradigm is mostly used in humans, as it recruits higher-level neural substrates for cognitive control (Monsell, 2003; Sohn et al., 2000).

Finally, the assessment of inhibitory control over prepotent or habitual responses is another category of tasks used to evaluate cognitive flexibility. An example of these paradigms is the stop-signal reaction task (SSRT). Subjects train to respond to a set of stimuli but then

are required to stop following a signal, which assesses action inhibition or the ability to flexibly inhibit a pre-planned physical response (Eagle et al., 2008). The Go/No-go task allows for measuring waiting impulsivity and inhibitory control. The paradigm involves two-choice discrimination with the presentation of a series of stimuli accompanied with "go" cues, which signal the need to respond to the stimulus, or with "no-go" cues, which require not to respond to the stimulus. If the frequency of go cues is larger than no-go cues, subjects might develop a prepotent tendency to respond, which must be inhibited when no-go cues are presented. Similarly, if conditions reverse, subjects must inhibit previous responses and develop new strategies (Costantini and Hoving, 1973; Mishkin and Pribram, 1955; Syed et al., 2015).

1.3 Psychological substrates of reversal learning

Predicting when and where a reward might occur allows humans and other animals to initiate and adapt their responses to optimise the number of rewards received (O'Doherty, 2011). Several learning and behavioural processes are required for optimal reversal learning performance.

1.3.1 Conditioning

Classical conditioning

The study of classical conditioning has its origins in the 19th century with Ivan Pavlov, who studied how animals, including humans, learn by association. Classical or Pavlovian conditioning refers to learning to associate an unconditioned stimulus (US) that elicits a biological response (unconditioned responses; UR) with a previously neutral stimulus that becomes a conditioned stimulus (CS). As result of learning this association, the CS alone comes to elicit the behavioural response originally associated with the US – the conditioned response (CR). This form of learning enables the subject to predict outcomes and exhibit preparatory or anticipatory behaviours. Importantly, in a Pavlovian conditioned procedure, there is no causal association between the animal's responses and the environmental outcome, but learning originates from repeated presentation (Mackintosh, 1974). The most common

association during this conditioning is the stimulus-response association (S-R), in which the CS is directly associated with the UR.

Instrumental conditioning

Instrumental or operant conditioning was first described by Burrhus Frederic Skinner as *"any and every voluntary behaviour that acts upon the environment to create a response"* (Skinner, 1938). This means that subjects exert control over events through a causal and direct link between their actions and subsequent outcomes. Two associative processes underlie instrumental behaviour: goal-directed actions and habitual behaviour.

Goal-directed actions are based upon knowledge of the contingency between actions and outcomes, so-called A-O associations, and the incentive value of those outcomes (Balleine and Dickinson, 1998; Cardinal and Everitt, 2004). Goal-directed processes dominate instrumental learning and behaviour in early stages of training, and are relatively flexible according to the needs fulfilled by particular outcomes.

As training progresses, actions become reflexive and less flexible in nature, eventually coming under the control of habitual processes. Habitual behaviours are based on S-R associations, which are sensitive to contiguous pairing of a specific action and reinforcer, as opposed to a causal relationship (Balleine and Dickinson, 1998). In opposition to Pavlovian S-R associations, instrumental S-Rs are created through positive or negative reinforcement, rather than associative learning.

Given the existence of multiple strategies to control behaviour, the question arises for their existence. In other words, if humans and other animals can behave in a goal-directed manner, why are less flexible behaviours required? The most compelling explanation is that each strategy offers a different trade-off between accuracy, speed, experience, and efficiency. Goal-directed behaviours guide the subject towards reliably achieving a goal, but flexibility is cognitively demanding and its implementation is relatively slow. In contrast, stimulus-driven behaviours may sometimes fail to meet an organism's current needs, but can be deployed quickly and require less computational resources (O'Doherty, 2011). Thus, a switch between strategies allows for the selection of the most economical strategy.

1.3.2 Discrimination learning processes

Solving a discrimination task requires the subjects to learn the value of each choice, forming accurate A-O associations (Izquierdo, 2017). However, during cognitive flexibility tasks such as reversal learning, the previous positive contingencies become negative, and *vice versa*. Successful reversal performance requires adjusting behaviour based on the representations of A-O contingencies. Following reversal, the stimulus associated with the reward (CS+), becomes associated with the lack of reward (CS-). To overcome this shift, a subject must stop responding to the original CS+, now CS-, a process opposed by perseverance. In contrast, the original CS- becomes the CS+, thus requiring the subject to start responding to the original CS-, a process opposed by learned non-reward.

Perseverative behaviour

Perseveration refers to the inappropriate repetition or maintenance of an action despite the absence or cessation of the original stimulus or outcome; or in cognitive terms, due to an incorrect abstract relationship between stimuli and goals (Garner et al., 2006). Perseverating on a specific action – linked to habitual behaviour – might be beneficial to achieve efficient and rapid responding. However, excessive perseverance is also maladaptive and underlies neuropathological disorders such as OCD (Serpell et al., 2009).

Typically, perseveration is framed as the inability to overcome reinforced associations, which in cognitive flexibility tasks corresponds to excessive approach to a previous CS- (Nilsson et al., 2015). To assess perseveration in reversal learning, the most common approach is to classify incorrect responses after stimuli reversal into early and late errors. It is widely assumed that the number of incorrect responses in early stages reflects the strength and stability of the old association and can therefore is an index of perseverative behaviour (Jones and Mishkin, 1972).

Learned non-reward

Learned non-reward emerges from the association of the CS with ‘no US’ (i.e. no reward). Following a contingency shift, learned non-reward interferes with the formation of a new strategy to optimise rewarded outcomes. An increased avoidance towards the original CS-

limits exploration of rewarding possibilities. In two-choice reversal learning tasks, the inability to overcome a non-rewarded association has also been named learned avoidance (Clarke et al., 2007) or learned irrelevance (Boulougouris et al., 2008).

Research has also focused on how learning is shaped by reinforced and non-reinforced outcomes, which ultimately guide responding. A growing number of studies has confirmed that some neurological disorders show an altered sensitivity to positive or negative feedback, which alter how learning is shaped. For example, one of the most prominent distortions in patients with major depressive disorder is abnormal sensitivity to negative feedback (Beats et al., 1996; Elliott et al., 1997; Rygula et al., 2018). Suffering from a bias towards negative feedback negatively skews the way that environmental information is processed (Elliott et al., 1997; Hales et al., 2014). The negative perception may lead to increased attention to received negative feedback and impair subjects' ability to perform in tasks that require cognitive flexibility, as well as impairing their wellbeing and daily experience (Elliott et al., 1997). The neural mechanisms underlying such impairment are unclear. Hence, understanding the mechanisms by which relevant cognitive biases are perceived and maintained is an important goal in the context of novel treatment development.

1.4 Neuroanatomical basis: the striatum

The basal ganglia are fundamental information processors in the mammalian brain. They consist of a group of subcortical forebrain and midbrain nuclei implicated in a broad range of behaviours, including movement, action selection, and reinforcement learning (Humphries and Prescott, 2010).

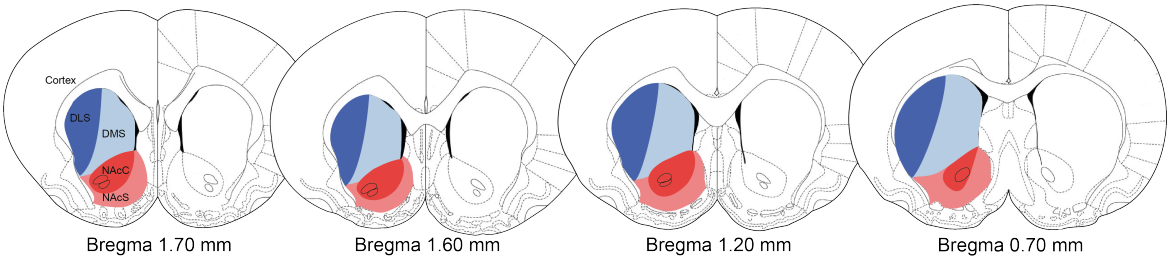
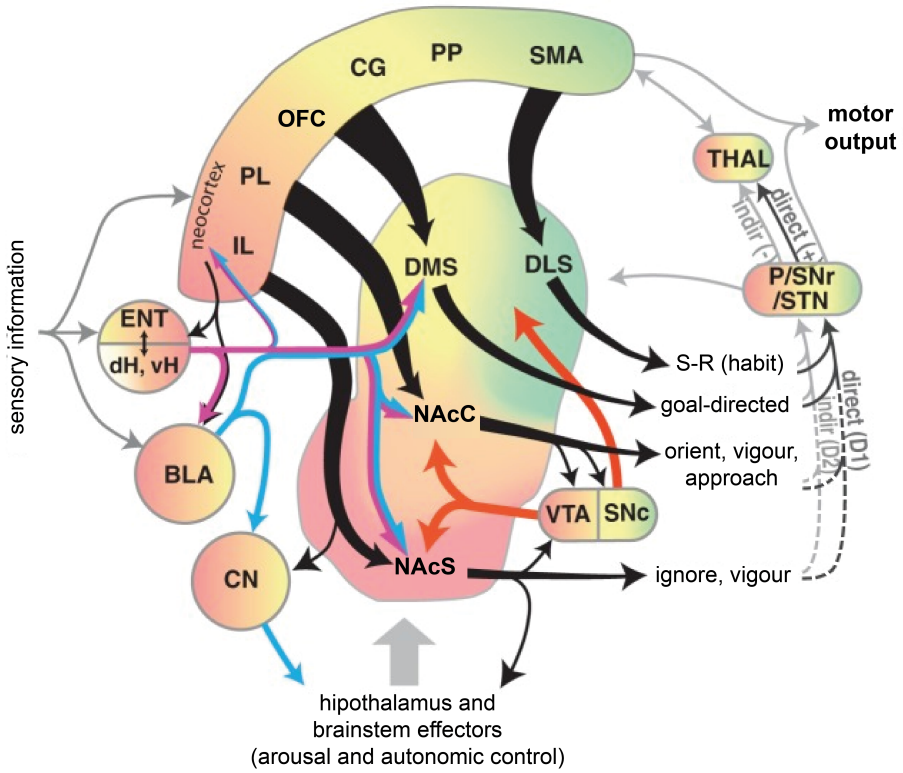


Fig. 1.2 Illustration of striatal subregions on coronal sections at 1.70, 1.60, 1.20 and 0.70 mm from Bregma. Abbreviations: dorsomedial (DMS), dorsolateral striatum (DLS), nucleus accumbens core (NAcC), nucleus accumbens shell (NAcS). Adapted from Paxinos and Watson (1998).



associations	[S, C] - O	[S, C] - [R, O]	S-R
association description	Pavlovian	model-based	model-free
effector coordinate system	autonomic	allocentric	egocentric
effector domain	somatic	cognitive/skeletal	skeletal
relative association speed	rapid	fast	gradual
response domain	emotive/motivation	goals	habits
response type	orient, approach, vigor	action by inference	typical response

Fig. 1.3 Diagram of the rat cortico-striatal circuit and its role in behavioural control. The schematic illustrates the proposed role of the dorsolateral striatum (DLS) in habitual behaviour, the dorsomedial striatum (DMS) in goal-directed actions, and the nucleus accumbens (NAc) in attributing vigour and directing orientation, approach or avoidance responses. Colour gradient indicates the gradient of afferent projections and topography (Voorn et al., 2004). Tapered arrows represent input convergence. Dorsostriatal output divides into the direct D1R-expressing pathway (with a disinhibitory effect, +), and the indirect D2R-expressing pathway (with an inhibitory effect, -). Ventrostriatal output also consists on D1R- and D2R-expressing MSNs, but their segregation between direct and indirect pathways is less categorical (dash line). Projections from the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc) are mainly dopaminergic and represented with red arrows. The table displays characteristic features of the different control systems of behaviour. Table abbreviations: stimulus (S), context (C), outcome (O), response (R). Circuit abbreviations: nucleus accumbens shell (NAcS), nucleus accumbens core (NAcC), substantia nigra pars reticulata (SNr), dorsal hippocampus (dH), ventral hippocampus (vH), pallidum (P), entorhinal cortex (ENT), subthalamic nucleus (STN), basolateral amygdala (BLA), central nucleus of the amygdala (CN). Neocortex abbreviations: infralimbic cortex (IL), prelimbic cortex (PL), orbitofrontal cortex (OFC), parietal cortex (PP), cingulate gyrus (CG), sensorimotor cortex (SMA). Adapted from Gruber and McDonald (2012).

1.4.1 Anatomical heterogeneity

The striatum in rodents is an anatomical and functional heterogeneous structure, which is typically subdivided into the dorsal and the ventral striatum (Fig. 1.2). The dorsal striatum can be further subdivided into the dorsomedial (DMS) and dorsolateral striatum (DLS). The ventral striatum mainly consists of the nucleus accumbens (NAc), which can be anatomically segregated into the NAc core (NAcC) and the NAc shell (NAcS) based on inputs and immunohistochemical markers (Zahm, 1999).

The striatum receives major inputs from glutamatergic and dopaminergic neurons. Glutamatergic inputs mainly arise from cortical regions, as well as thalamic and limbic regions. Cortical efferents innervate the striatum following a dorsomedial-ventrolateral topographical organisation (for review see (Haber, 2016)). The DMS receives inputs from associative regions, including projections from more dorsal regions of medial prefrontal cortex (mPFC), orbitofrontal cortex (OFC), primary motor and somatosensory cortices, as well as amygdala, thalamus, and DA midbrain systems, including the substantia nigra pars compacta (SNc) (Wall et al., 2013) (Fig. 1.3). Conversely, the DLS is innervated by sensorimotor cortices, thalamus, and DA midbrain systems (Burke et al., 2017). Conversely, the NAc receives projections from the prefrontal cortex (PFC), including the OFC, amygdala, thalamus, midbrain DA system [mainly, ventral tegmental area (VTA)], and the laterodorsal tegmentum (Berendse et al., 1992; Berendse and Groenewegen, 1990; Boeijinga et al., 1993; Ikemoto, 2007). The NAcS and NAcC receive inputs from the infralimbic cortex (IL) and the prelimbic cortex (PL) (Keistler et al., 2015; Sesack et al., 1989; Vertes, 2004). Dopaminergic innervation into the NAc has a mediolateral topography within the NAcC and NAcS. DA neurons in the posteromedial VTA generally project to the ventromedial striatum, including the NAcS, whereas anteromedial VTA efferents largely project to the ventrolateral striatum, including the NAcC, but also in part the NAcS (Ikemoto, 2007).

The NAcC and NAcS have other defining features that further enable their differentiation. In rats, the NAcC has a higher cell density (Meredith et al., 1992) and higher DA and 5-HT metabolism (Deutch and Cameron, 1992) compared with the NAcS. In addition, there is recognised variation in the density of afferents and efferents in the two NAc subregions (Salgado and Kaplitt, 2015), with the NAcS receiving a denser input of glutamatergic inputs than the NAcC (Mingote et al., 2019).

1.4.2 Cellular heterogeneity

Striatal neurons can be classified into two main subgroups: medium spiny neurons (MSNs) and interneurons. MSNs are the primary cell type in both dorsal and ventral regions, making up over 90% of total striatal neurons. These are GABAergic (γ-aminobutyric acid) inhibitory neurons that receive excitatory cortical efferents (Tepper and Bolam, 2004). MSNs also receive dopaminergic modulatory input from the SNc or the VTA *via* the nigrostriatal and mesolimbic loops. This dopaminergic input can have different modulatory effects depending on the targeted MSNs pathway: direct and indirect pathways.

The direct and indirect pathways are the two main channels of information flow through the basal ganglia, and are constituted by two principal types of MSNs (Gerfen et al., 1990) (Fig. 1.4). The direct pathway originates from a subpopulation of MSNs expressing DA D1 receptors (D1R), which are Gs-coupled receptors. These receptors increase neuronal activity by activating adenylate cyclase signalling that catalyses the conversion of cytosolic adenosine triphosphate (ATP) into cyclic-adenosine monophosphate (cAMP). D1R expressing neurons project directly from the striatum to the substantia nigra pars reticulata (SNr), also sending projections to the internal part of the globus pallidus (GPi). The indirect pathway originates from the subpopulation of MSNs expressing dopamine D2 receptors (D2R), which are Gi-coupled receptors. These receptors reduce cell activity by inhibiting adenylate cyclase signalling *via* their G-coupled protein. Neurons belonging to the indirect pathway indirectly innervate the SNr by projecting from the striatum to the external part of the globus pallidus (GPe). GPe neurons release GABA to the subthalamic nucleus (STN), and finally the STN sends glutamatergic projections to the SNr. These output nuclei modulate the thalamus, which in turn sends input to the cortex, closing the cortico-striato-thalamo-cortical loop (for review see (Haber, 2016)) (Fig. 1.4).

Direct and indirect striatal MSNs also express additional distinctive opiate peptides. The direct pathway, expressing D1R, also contains substance P and dynorphin, whereas the indirect pathway, expressing D2R, co-expresses enkephalin. These neuropeptides are hypothesised to modulate dopaminergic input to the striatum (Steiner and Gerfen, 1998).

Although the segregated view of striatal outputs is widely accepted, it is also an over-simplification. For example, extensive reciprocal connections exist within the basal ganglia, including projections from the globus pallidus (GP) back to the striatum (Bevan et al., 1998). Within the same striatum, MSNs send dense axon collaterals to other MSNs modulating their

activity (Burke et al., 2017; Tunstall et al., 2002). Hence, MSNs belonging to the direct and indirect pathways can influence the output from the other pathway, following the so-called process of "lateral inhibition". Recently, the division of MSNs DA receptors subtypes has been questioned, since it might not involve a complete segregation (Cazorla et al., 2014). The dichotomy of direct and indirect pathways also seems to hold more for the dorsal, but less so for the ventral striatum, where the proportion of MSNs expressing both D1R and D2R is higher (Bertran-Gonzalez et al., 2008; Matamales et al., 2009), and D1R- and D2R-MSN projection targets are less segregated (Kupchik et al., 2015).

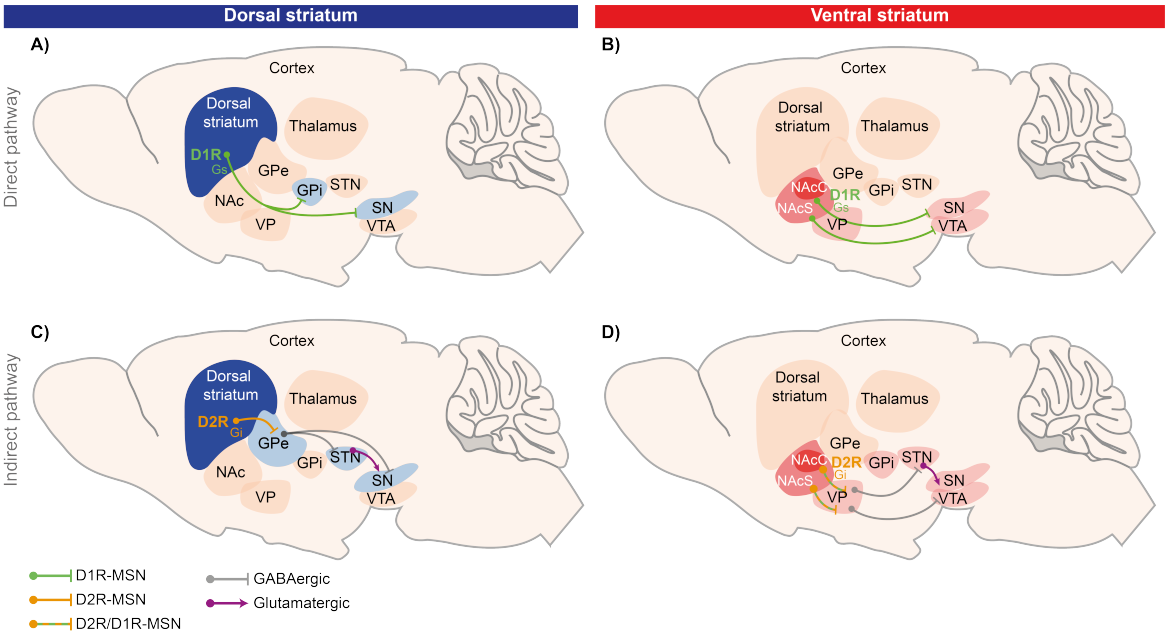


Fig. 1.4 Direct and indirect pathways in the dorsal and ventral striatum. A) Direct pathway in the dorsal striatum expresses D1R, coupled to Gs protein, and project to the globus pallidus pars interna (GPi) and substantia nigra (SN). B) Direct pathway in the ventral striatum expresses D1R and predominantly projects to the SN from the nucleus accumbens core (NAcC), and ventral tegmental area (VTA) from the nucleus accumbens shell (NAcS). C) Indirect pathway in the dorsal striatum expresses D2R and projects to the globus pallidus pars externa (GPe), which projects directly or *via* the subthalamic nucleus (STN) to SN. D) Indirect pathway in the ventral striatum expresses both D1R and D2R, and projects from the NAcC to the dorsolateral part of the ventral pallidum (VP), which projects to the STN and reaches the SN; or from the NAcS to the ventromedial part of the VP, sending GABAergic projections to the VTA. Adapted from Soares-Cunha et al. (2016).

1.4.3 Role of the striatum in learning

Anatomic segregations also lead to functional dissociations. In general terms, the dorsal striatum plays a role in motor planning, action selection and S-R habit learning, whereas the NAc is involved in regulating motivated behaviour and reward-related learning (Isomura et al., 2013), although these functions can also overlap. Thus, the stronger input from the limbic cortex and VTA, rather than SNc, into the NAc makes it the region where information about reward and motivational drive are integrated to guide behavioural performance (Mogenson et al., 1980), as opposed to the dorsal striatum's role in sensorimotor integration (Robbins and Everitt, 1992).

The dorsal striatum has been suggested to play a role in cognitive flexibility due to its implication in goal-directed actions and habitual behaviours. Specifically, the DLS mediates habitual, whereas the DMS regulates goal-directed behaviours (Brigman et al., 2013; Corbit et al., 2014; Yin and Knowlton, 2006). Yin and colleagues found that lesions of the DLS disrupt habit formation in instrumental learning, whereas outcome expectancy is preserved (Yin et al., 2004). They also observed that DLS lesions impair performance in an outcome devaluation task, which is a commonly used task to assess habitual behaviours. Conversely, lesions to the DMS had no effect on behaviour in this paradigm (Yin et al., 2004). Similarly, DLS DA-depleted rats became sensitive to reward-devaluation and incapable of forming S-R habits in an operant conditioning task (Faure et al., 2005).

The DMS is necessary for the acquisition and expression of A-O instrumental learning. Lesions of the posterior part of the DMS during pre-training stages blunt sensitivity to both contingency and outcome degradation (Yin et al., 2005). It has also been observed that lesions of the DMS before and after training impair sensitivity to devaluation and degradation contingencies, which highlights the importance of the DMS (especially its posterior part) in both acquisition and expression of A-O associations (Yin et al., 2005). In addition, *in-vivo* recordings during behavioural shifting showed that the DMS (together with the OFC) becomes more engaged, and the DLS less engaged, when behaviour shifts to goal-directed responding, showing that both regions dynamically encode the shift between goal-directed actions and habitual behaviours (Gremel and Costa, 2013).

Conversely, the NAc does not seem to be *required* for goal-directed or habitual behaviours, since excitotoxic lesions of this structure leave performance intact, but it does appear to

regulate these behaviours by integrating limbic information (Cardinal et al., 2002); for example, in processing reward-related stimuli.

Human imaging studies show that both the dorsal and ventral striatum are recruited during reversal-learning tasks, and lesions in the basal ganglia impair performance in this form of cognitive flexibility (Cools et al., 2002; Rogers et al., 2000). Lesions in the DLS impair late phase of reversal learning (Brigman et al., 2013). Similarly, DLS lesions disrupt habit formation in instrumental learning, while maintaining outcome anticipation (Yin et al., 2004). In contrast, lesions of the DMS in rats impair various forms of reversal, while leaving unaffected the retention of the initial discrimination (Clarke et al., 2008; Ragozzino and Choi, 2004). This effect was suggested to be due to a failure in suppressing perseverative responses to previous reward contingencies (Castañé et al., 2010). Indeed, blocking the DMS in marmosets produces perseverative behaviour, leading to impaired reversal-learning performance, in a similar manner to OFC lesions (Clarke et al., 2014, 2008).

In contrast to the dorsal striatum, the role of the ventral striatum in cognitive flexibility is controversial. Lesions in this region did not alter reversal learning, latency to collect the reward, number of omissions or locomotor activity (Castañé et al., 2010), but the ability for set-shifting tasks was affected in rats (Floresco et al., 2009). It has been suggested that the NAc is more involved in complex forms of behavioural flexibility involving changes in strategies, but it does not appear to make a critical contribution to simpler forms of these processes, such as visual discrimination or reversal learning (Castañé et al., 2010; Floresco et al., 2009; Kehagia et al., 2010). Nevertheless, lesions in the NAcS in rats impaired performance in a probabilistic reversal learning task, while lesions in the NAcC affected reward collection latencies but not overall performance (Dalton et al., 2014).

1.5 Neurochemical basis: dopamine

DA, or 3-hydroxytyramine, was discovered over 60 years ago as the precursor of noradrenaline (Montagu, 1957). Shortly thereafter, Carlsson and colleagues showed that DA concentration in the striatum was higher than in the rest of the brain, even if its concentration of noradrenaline was low, indicating that DA was not only a precursor, but also a neurotransmitter in its own right (Carlsson et al., 1958). Subsequently, DA has been intensively investigated due to its demonstrated role in a multitude of brain functions, including learning,

memory, motivation, and emotional behaviours, which has had a significant impact in the applied fields of neurology, psychiatry and psychopharmacology (Iversen and Iversen, 2007).

1.5.1 Dopaminergic nuclei and projections

DA, together with adrenaline and noradrenaline, constitute the monoamine neurotransmitter family of the catecholamines, which are organic compounds with a catechol and a side-chain amine. Catecholamines are synthesised from the amino acid l-tyrosine and different enzymes catalyse its transformation into each of the above neurotransmitters. The primary cytoplasmatic synthetic pathway of DA involves the transformation of l-tyrosine into L-DOPA *via* tyrosine hydroxylase (TH) and L-DOPA into DA *via* DOPA decarboxylase. Extracellular DA can be reuptaken by the DA transporter (DAT). After its cytoplasmatic synthesis or reuptake, DA is accumulated into vesicles *via* the vesicular monoamine transporter 2 (VMAT2) (Fig. 1.7). These enzymes may work as neuromarkers for DA neurons. However, molecular heterogeneity exists between different DA cell subpopulations (Morales and Margolis, 2017). While various cellular compartments of DA neurons express high levels of TH – and this is therefore often used to detect DA neurons –, some TH positive neurons lack expression of VMAT2 or DAT or co-release other neurotransmitters, like glutamate (Li et al., 2012; Morales and Margolis, 2017).

The neuronal soma that produce DA are localised in a small number of regions. In the mammalian brain, there are nine DA-producing nuclei, from A8 to A16 (Fig. 1.5; (Björklund and Dunnett, 2007). From these, three are located in the midbrain – A8, A9 and A10 –, with the A9 (SNc) and A10 (VTA) receiving the most attention due to their role in reinforcement learning and reward processing (Tritsch and Sabatini, 2012).

The three major dopaminergic pathways are the nigrostriatal, mesolimbic, and the mesocortical (Fig.1.5). The tuberoinfudibular pathway is a fourth system that merges from the arcuate or infundibular nucleus in the tuberal region of the hypothalamus and innervates the median eminence, attached to the infundibulum. Unlike the other pathways, it modulates the release of hormones in the blood, such as prolactin (Weiner and Ganong, 1978), but will not be discussed further in this thesis.

Although this standard segregation of these pathways is well accepted, it is an over simplification. Neurons from the SNc also send axons to the limbic system and the cortex.

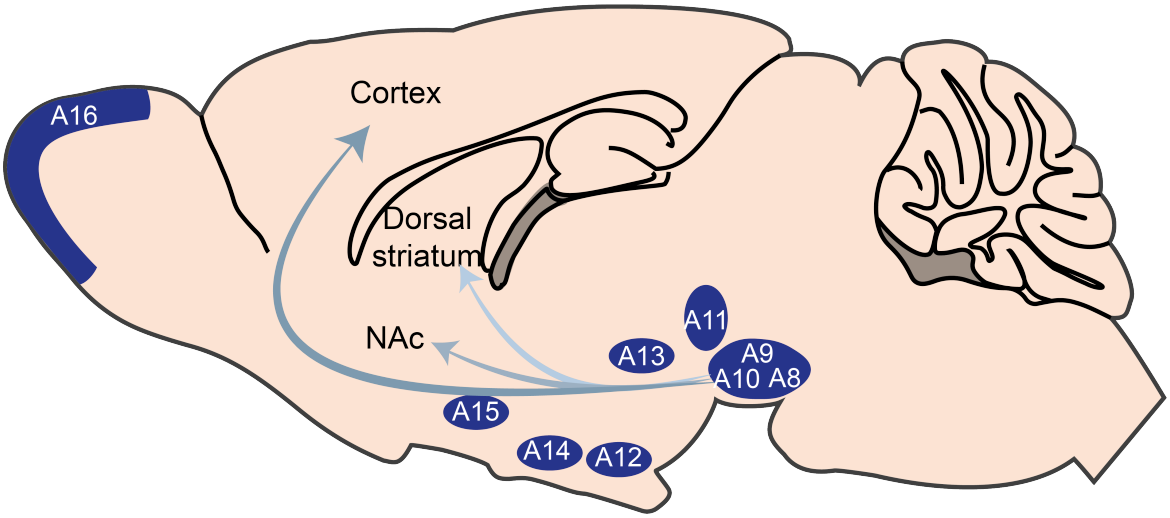


Fig. 1.5 Dopaminergic nuclei and projections of interest of the rat brain in sagittal view. Nine major cell groups are represented in navy blue. Projections of interested are also illustrated: nigrostriatal (light blue), mesolimbic or mesoaccumbal (light grey-blue), and mesocortical (grey-blue). Adapted from Björklund and Dunnett (2007).

In addition, despite these pathways being anatomically and functionally different, their cell bodies are intermingled in the VTA and SN (Björklund and Dunnett, 2007). Within the basal ganglia, it is not only the striatum that receives dopaminergic inputs form the midbrain – which processes DA signalling *via* the direct and indirect pathways discussed above -, but also the GP, ventral pallidum and STN (Hassani et al., 1997; Lindvall and Björklund, 1979). Thus, downstream effects of basal ganglia signalling extend further than the dorsoventral striatal influence (or caudate-putamen in primates), allowing for a more complex and comprehensive influence on the regulation of the circuitry. See below for further details on the striatal dopaminergic network modulating cognitive flexibility.

1.5.2 Dopamine spike firing and release

DA neurons exhibit two distinct modes of firing activity: tonic and phasic (Goto et al., 2007; Grace, 1991). Understanding the specific functions that DA mediates depends fundamentally on phasic *versus* tonic activity, which governs DA release dynamics and synaptic sites of action.

Tonic firing refers to slow and spontaneous firing that is driven by changes in membrane conductance mediated by glutamatergic inputs (Grace and Bunney, 1984; Grace and Onn, 1989) and GABAergic inhibition (Grace and Bunney, 1979). Importantly, when DA neurons are tonically activated, release of DA exceeds the synaptic cleft and spills into the extracellular space. Tonic extracellular DA levels depend on the number of DA neurons that spontaneously spike in a tonic manner (Grace, 1991) and results in small concentrations varying over time e.g. 4-20 nM within the striatum (Keefe et al., 1993; Parsons and Justice, 1992). Although these concentrations are too low to have an effect on post-synaptic neurons, they are sufficiently high to modulate the activity of pre-synaptic receptors (i.e. autoreceptors) that regulate release of DA into the synaptic space (Grace, 1991, 2000). Thus, elevated tonic firing can attenuate phasic DA signalling (Grace, 1991).

Phasic firing, in contrast, refers to sharp changes in firing rate – bursting activity – leading to large changes in DA release. DA release under these conditions depends on glutamatergic excitatory innervations from a number of brain regions, including the STN and the pedunculopontine tegmentum (Floresco et al., 2003; Futami et al., 1995; Smith and Grace, 1992). Phasic DA release is of high amplitude (e.g. hundreds of μM to nM), transient and powerful (Goto et al., 2007; Grace, 1991). However, DA is quickly removed from the synaptic cleft *via* high affinity uptake that limits the diffusion of DA into the extracellular space. Phasic signalling is proposed to be the primarily involved in reinforcement learning (Schultz et al., 1997), which is discussed in more detail below.

1.5.3 Dopamine as a teaching signal

Reward prediction errors

Seminal work from Schultz and colleagues (Schultz et al., 1997) provided insight into the mechanisms that promote learning resultant from reward prediction errors (RPEs). RPEs refer to the contrast between learnt expectations and actual outcomes, and are used to update current beliefs and adjust behaviour (Fig. 1.6). Using electrophysiological recordings in the midbrain of macaque monkeys, Schultz and colleagues demonstrated that in a task were a cue would predict the delivery of reward, DA firing rate would increase in response to an unpredicted reward. If animals were over-trained, they would learn to associate a CS as predictor of reward, and DA neurons would shift from firing upon presentation of the

reward to firing upon presentation of the cue. In addition, receiving a larger reward than expected would transiently increase dopaminergic firing rate (i.e. induce a burst), causing a so-called positive RPE. In contrast, a smaller reward would pause tonic firing (i.e. induce a dip), causing a negative RPE. If the reward did not differ from expectation, DA signalling would remain unchanged (Schultz, 2013; Schultz et al., 1997). Therefore, DA is involved in reinforcement learning, especially when expectations are violated.

Since the proposal that midbrain DA neurons signalling encodes RPEs was first suggested, it has been widely confirmed using different behavioural paradigms (Tobler et al., 2003; Waelti et al., 2001), rewards with distinct properties or dimensions (Lak et al., 2014), and across species (Cohen et al., 2012; O'Doherty et al., 2006; Takahashi et al., 2016). Further, a causal link between neuronally encoded RPEs and reinforcement learning has recently been reported. Thus, it has been shown that timed inhibition of DA activity disrupts learning (Hamid et al., 2016) whereas hyper-activation of DA neurons with *in-vivo* optogenetics to generate an artificial positive RPE increases cue-driven reward-seeking behaviour (Steinberg et al., 2013). In contrast, inhibition of DA neuronal activity, that simulates an artificial negative RPE, is sufficient to impair associative learning (Chang et al., 2015). In agreement with these results, DA transmission also provides a prediction error signal during reversal learning, transiently decreasing in response to errors following a shift in response-outcome contingencies, and increasing after unexpected rewards when animals start interacting with the previously unrewarded stimulus (Klanker et al., 2015; Verharen et al., 2018).

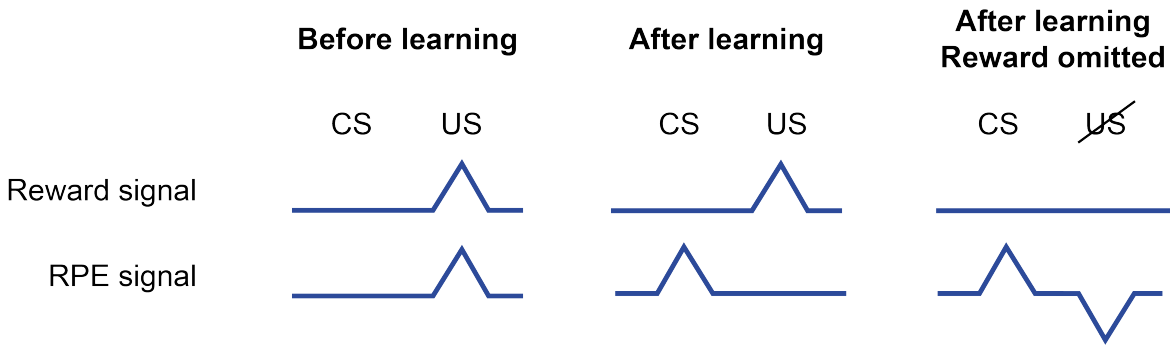


Fig. 1.6 Illustration of reward and reward prediction error (RPE) signals. Before learning, dopaminergic neurons signal a positive RPE in the presence of an unconditioned stimulus (US; reward). After learning, the conditioned stimulus (CS; e.g. cue) predicts the delivery of reward and dopaminergic neurons signal a positive RPE paired with the CS, not the US. When reward is omitted after learning, activity of dopaminergic neurons is depressed at the time when the US was expected (i.e. a negative RPE). Adapted from Verharen (2018).

Model of the basal ganglia

As previously discussed, midbrain DA neurons send dense projections to the striatum, where RPE signals are processed *via* the direct and indirect pathways, expressing D1R and D2R, respectively (see section 1.4.2). Due to their G-coupled protein, post-synaptic D1R are sensitive to bursts in DA signalling. Thus, positive associations leading to positive RPEs are strengthened *via* D1R within the direct pathway. In contrast, post-synaptic D2R are sensitive to transient dips in DA transmission, so that negative associations leading to negative RPEs might be weakened *via* D2R within the indirect pathway (Frank et al., 2004; Yapo et al., 2017).

Frank and colleagues' proposed a model for this segregation. In seminal work, they observed that reinforcement learning was altered in Parkinson's patients only after they had taken their dopaminergic medication (Frank et al., 2004). They tested patients using a probabilistic selection task based on a two-choice visual discrimination, which could be solved either by learning to approach the positive stimulus, or to avoid the negative stimulus. They found that patients had a selective decline in learning from losses. A proposed explanation was that over-physiological states of DA hindered learning from negative feedback by leaving D2R-expressing striatopallidal neurons insensitive to dips in DA, thereby blocking learning from negative feedback. In contrast, diminished levels of DA would hypothetically prevent D1R-expressing nigrostriatal cells from detecting DA burst firing, thus blunting learning from positive feedback (Cox et al., 2015; Frank et al., 2004).

Consequently, to identify the specific role of different DA pathways in reversal learning, it is important to differentiate between learning from positive and negative feedback, or behaviours approaching the positive stimulus from behaviours avoiding the negative stimulus.

1.5.4 Dopaminergic receptors

Receptors subtypes

The existence of multiple types of DA receptors was first proposed in 1976 (Cools and Van Rossum, 1976). At least five distinct subtypes of DA receptors, D1R, D2R, D3, D4 and D5 receptors (D3R; D4R; D5R) have been described (Surmeier et al., 2007, 1996) and classified into two families: D1-like receptor, encompassing D1R and D5R; and D2-like receptors,

encompassing D2R, D3R and D4R. The classification was established based on the genomic organisation of DA receptors, but primarily on the G protein they are coupled to: D1-like are Gs-coupled, leading to activation of the neuron, while D2-like receptors are Gi-coupled, which inhibits neuron, as reviewed in section 2.2.2. D2R can be subdivided based on two distinct isoforms that arise from alternative gene splicing: D2R short (D2S) and long (D2L) variants (Giros et al., 1989), which differ in their localisation and function (Centonze et al., 2004).

The receptors subtypes also differ in their baseline affinity states. Affinity is defined as the strength of the intermolecular force between the receptor and its ligand (in this case, DA). For high-affinity receptors, the ligand takes longer to dissociate from the receptor than for low-affinity receptors. This affects the concentration of ligand required to saturate occupation of a receptor: low-affinity receptors require larger concentrations than high-affinity receptors. D2-like receptors have greater affinity than D1-like receptors (Richfield et al., 1989). Hence, D2-like receptors are considered to be more sensitive to low levels of DA, while D1-like receptors are more easily activated by phasic DA changes (Richfield et al., 1989) – although the high-affinity of D2-like receptors has recently been challenged (Yapo et al., 2017). This differentiation has given rise to behavioural models linking D1-like receptors with reward learning and D2-like receptors with overall motivational levels (Missale et al., 1998). However, recent studies show that D2-like receptors are sensitive to pauses and phasic increases in DA neuron firing and release (Marcott et al., 2014).

Within the central nervous system (CNS), D1-like receptors are present in higher density in the striatum, SNr, and olfactory bulb (De Keyser et al., 1988). A lower level of expression has been reported in the entopeduncular nucleus, cerebral aqueduct, ventricles, with even lower expression in the dorsolateral PFC, cingulate cortex and hippocampus (Boyson et al., 1986; Savasta et al., 1986). D2-like receptors are mainly expressed in the striatum. They are also present in the GPe, cerebral cortex, amygdala, and pituitary gland (De Keyser et al., 1988; Jackson and Westlind-Danielsson, 1994).

Functional interactions of DA receptors in the striatum

Within the striatum, D1R and D2R are found throughout the region in elevated density, from dorsal to ventral subregions (Gerfen et al., 1990; Matamales et al., 2009; Surmeier et al., 2009). D5R are present in low levels in MSNs, but are highly expressed in interneurons

(Rivera et al., 2002). D3R and D4R are expressed in low levels in the dorsal striatum (Mrzljak et al., 1996; Surmeier et al., 1996). There is also significant expression of D3R, but not D4R, in the NAc (Mrzljak et al., 1996; Richtand et al., 1995).

Although the segregation of striatal MSNs between direct and indirect pathways has been widely accepted, the separate distribution of D1R and D2R within this anatomical dichotomy is less clear. Much work has been done to distinguish whether these receptors are expressed in different MSNs or co-expressed in the same neurons (Aizman et al., 2000; Surmeier et al., 1992, 1996). In the dorsal striatum, only a minority (i.e. 5%) of neurons co-express D1R and D2R and it is widely accepted that the D1R and D2R defined MSNs project to the SNr and GPe, respectively, following the direct and indirect pathway segregation (Surmeier et al., 1996). Neurons expressing D1R, D2R or co-expressing both account for the totality of MSNs. However, D1-like and D2-like receptors do not appear to be that well separated. For example, it has been reported that some D2R-expressing MSNs also contain D5R, and a high number of D1R-expressing MSNs (i.e. up to 70%) contain D3R and/or D4R. Nonetheless, as previously mentioned, D3R, D4R, and D5R are present in a much lower level than D1R and D2R (Surmeier et al., 1996).

In the ventral striatum, the dissociation is further complicated in relation to the dorsal striatum. On one hand, the MSN population co-expressing D1R and D2R is higher, being 6% in the NAcC and 17% in the NAcS (Bertran-Gonzalez et al., 2008; Matamales et al., 2009). On the other hand, a comparatively higher number of cells also express D3R. Double-staining studies have estimated that 33% of D2R-expressing MSN in the NAc (NAcC and NAcS) co-express D3R, compared to 22% in the NAcC and 54% in the NAcS (Le Moine and Bloch, 1996; Schwartz et al., 1998).

Less is known about post-synaptic receptors in interneurons. Cholinergic interneurons express D5R and D2R, but not D1R (Bertran-Gonzalez et al., 2008; Rivera et al., 2002). It is unknown if D3R and D4R are present in striatal interneurons. It has also been suggested that GABAergic terminals originating from interneurons express D2R (Bertran-Gonzalez et al., 2008).

Pre-synaptically, D2R-like receptors are expressed on glutamatergic (Hsu et al., 1995), GABAergic (Guzmán et al., 2003), cholinergic (Pisani et al., 2000), and dopaminergic (Benoit-Marand et al., 2001) axon terminals in the striatum. Some studies suggest that D2R are mainly expressed pre-synaptically, act as autoreceptors, and belong to the D2S isoform.

In contrast, D2R expressed in post-synaptic neurons belong to the D2L isoform (Usiello et al., 2000), although recent studies suggest post-synaptic D2R can also belong to the D2S isoform (Gantz et al., 2015). Exclusively in the NAc, D1-like receptors have also been observed pre-synaptically on glutamatergic terminals (Dumartin et al., 2007).

Despite the complexity of the distribution of DA receptors in the striatum, D1R- and D2R-expressing neurons present morphological and dendritic excitability differences, and D1R and D2R receptors represent the greatest population of expressed receptors, both in the dorsal and ventral striatum (Day et al., 2008; Gertler et al., 2008; Richtand et al., 1995). Therefore, they are still classified as D1R- and D2R-*dominant* MSNs. Consequently, from this point forward, D1-like and D2-like receptors will be referred to as D1R and D2R, respectively.

Interaction of D2R and adenosine receptors

Adenosine is considered to be a neuromodulator in the CNS, where it regulates neuronal excitability and neurotransmitters release *via* its four subtypes of G-protein-coupled receptors: A1, A2A, A2B and A3 (Borea et al., 2018). Whereas adenosine A2A (A2AR) and A2B receptors preferably interact with Gs proteins, A1 and A3 interact with Gi proteins.

Adenosine A1 receptors are present in nearly all brain areas and suppress neuronal excitability (Borea et al., 2018). In contrast, A2AR are almost exclusively found in DA-rich areas – such as the striatum –, where they promote neuronal activation (Borea et al., 2018; Vontell et al., 2010). Adenosine A2B and A3 receptors are mainly expressed in the peripheral, and have therefore received less attention in research on cognition (Daly et al., 1983; Dixon et al., 1996; Zhou et al., 1992).

One of the most important roles of adenosine is to induce a brake in the CNS. This inhibitory function is largely modulated by the activation of one subtype of receptors: A2AR, which are specifically expressed in striatopallidal neurons in the striatum (DeMet and Chicz-DeMet, 2002; Ferré, 2008) (Fig. 1.7). These neurons not only express the highest density of D2R in the brain, but also the highest density of A2AR of any other neuron type or brain area (Ferré, 2008). It is well documented that adenosine exerts its function in MSNs from the indirect pathway through intermolecular interactions between A2AR and D2R, which form receptor heteromers. This shared localisation, combined with evidence that A2AR and D2R

have an antagonistic inter-relationship, implicates adenosine in functions associated with DA signalling (Ferré, 2008; Furlong et al., 2017; Nunes et al., 2013; Santerre et al., 2012).

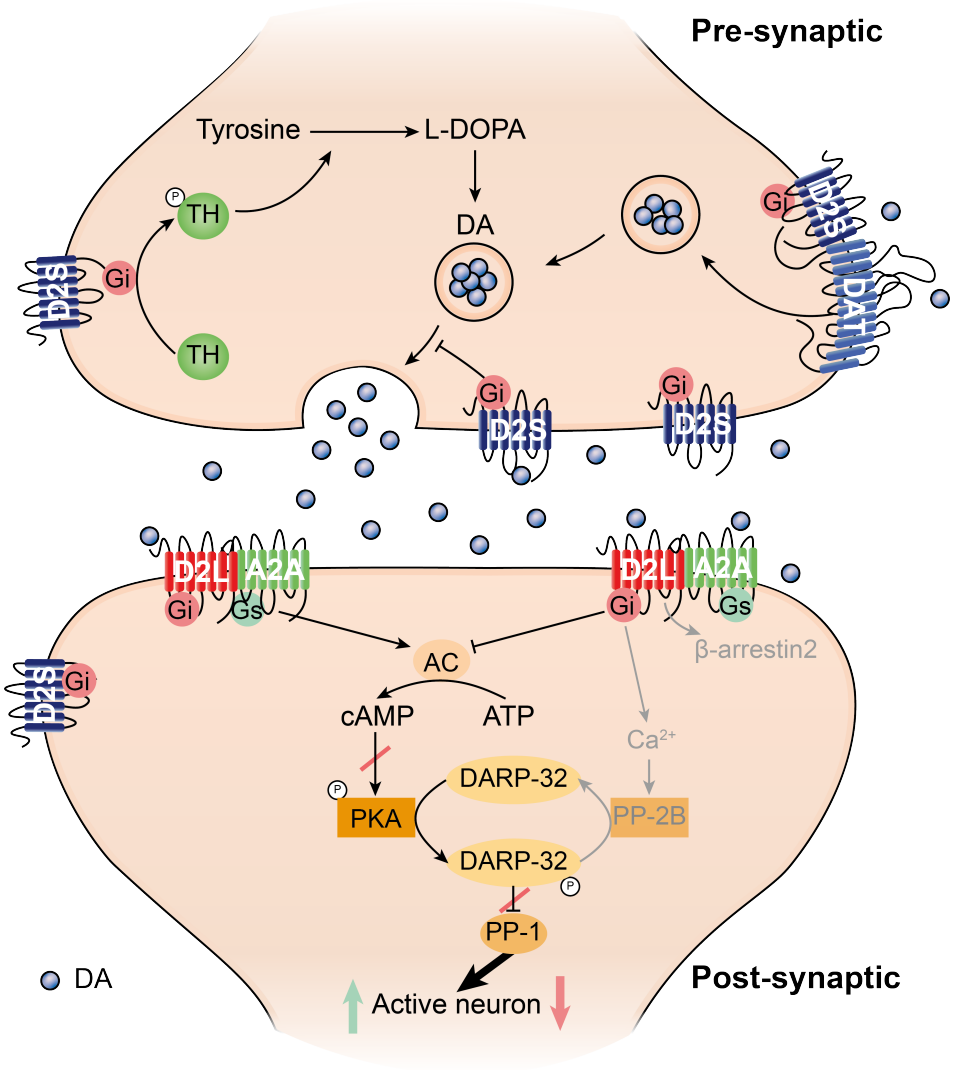


Fig. 1.7 Illustration of D2R-mediated pre- and post-synaptic signalling in neurons. At the pre-synaptic cell, D2R mainly belong to the short isoform (D2S), where they regulate synthesis and release of dopamine (DA). D2S regulates DA synthesis by phosphorylating (P) tyrosine hydroxylase (TH), which transforms tyrosine into L-DOPA, the DA precursor. Pre-synaptically, D2S also modulate activity of the DA transporter (DAT). At the post-synaptic cell, D2R mainly belong to the long isoform (D2L), although D2S can also be present. Post-synaptically, D2R form heteromers with adenosine 2A (A2A) receptors (A2AR). Whereas D2R are coupled to Gi protein and inhibit adenylyl cyclase (AC), A2AR are coupled to Gs and activate AC. Following the AC downstream cellular pathway, D2R activation leads to cell inhibition, whilst A2AR lead to cell activation. D2R can also act *via* calcium (Ca^{2+}) or G-independently *via* β-arrestin2.

1.5.5 Role of DA in reversal learning

DA regulates synaptic plasticity in brain areas noted for modulating reversal-learning performance: the cortex and the striatum (Cagniard et al., 2006; Calabresi et al., 2007), encodes RPEs (Schultz, 2013), and dopaminergic interventions affect different forms of learning (Steinberg et al., 2013). Striatal DA plays a key role in modulating reversal learning, as shown by induced non-perseverative impairments in reversal learning following DA depletion in the striatum in marmosets (Clarke et al., 2011, 2007).

Further research supports the central involvement of DA in reversal learning (Izquierdo et al., 2017). Increases in striatal DA levels following methylphenidate administration improve reversal learning in humans (Clatworthy et al., 2009). In rats, Klanker and colleagues showed that DA levels increase in the ventromedial striatum in response to unexpected rewards in a spatial reversal-learning task (Klanker et al., 2013), suggesting that striatal DA modulates its performance.

DA receptors subtypes have been differentially involved in mediating reversal learning. Many studies have investigated the potential role of each subtype in this form of behavioural flexibility. Systemic administration of a D1R agonist impaired early stages of visual reversal-learning task on touchscreens in mice (Izquierdo et al., 2006). It has been proposed that the circuit innervating the NAc from the basolateral amygdala mediates this effect, since it enhances behaviours related to reward *via* D1R (Stuber et al., 2010). Indeed, local infusions of a D1R antagonist into the NAcC improved performance during the early phase of visual serial reversal learning tested on touchscreens (Sala-Bayo et al., 2020).

Systemic administration of D2R agonists impaired spatial reversal learning in rodents, whereas D2R antagonists did not affect reversal performance (Boulougouris et al., 2009). Low DA availability is related to poor performance in rodents (Laughlin et al., 2011), non-human primates (Groman et al., 2011), and humans (Jocham et al., 2009). D2R antagonism, but not D1R, in the DMS and DLS impairs reversal learning performance at different stages in a complementary manner i.e. the manipulations in the DMS affect mid stages, whereas of the DLS affect overall performance, including early and late stages (Sala-Bayo et al., 2020). A cumulative piece of work indicates that reversal learning relies on optimal balance of D2R function. D2R functional balance has canonically been referred to as an inverted U-shaped function of DA, or a triphasic effect of D2R agents, showing that both low and high levels of DA function lead to cognitive impairments, and behavioural improvements may be found in

an optimal level in between both extremes in DA signalling (Horst et al., 2019; Yerkes and Dodson, 1908).

1.6 Thesis overview

1.6.1 Summary and aim

Cognitive flexibility refers to how individuals adapt their behaviour to changes in the environment. Although important for survival and wellbeing, cognitive flexibility is impaired in a wide range of neurological and neuropsychiatric disorders, including Parkinson's disease and OCD. Optimal flexibility is known to depend on DA neurotransmission in the CNS. During reversal, subjects must adapt and respond to the formerly non-rewarded stimulus whilst ignoring the initially rewarded stimulus. Learning on this task thus requires constant shifts in behaviour in response to positive (rewarded) and negative (non-rewarded) feedback.

Although the concepts of reward prediction, learning and flexible behaviour have been associated with DA, the precise mechanism and brain loci underlying the effects of DA on flexible decision-making remain unclear. Many unanswered questions reflect the present inability to tie these aspects of its function together, a likely consequence of an imprecise level of investigation at the brain and behavioural levels. These questions include: what brain subregions and receptors modulate cognitive flexibility? Do they modulate reversal learning in a complementary and dynamic manner? Which stages or cognitive subprocesses does DA modulate? How are positive or negative feedback signals integrated within the striatum? Do RPEs causally modulate reversal learning performance? If so, *via* what neuronal projections and in what cognitive subprocesses?

Thus, the main aim of my thesis is to provide more refined insights into the above questions, with the overarching hypothesis that DA modulates reversal learning performance by signalling positive and negative RPEs within the direct (rewarded) and indirect (non-rewarded) pathways, respectively.

1.6.2 Outline

In this thesis, cognitive flexibility was inferred in experimental rats by evaluating their performance in reversal-learning tasks involving a two-choice discrimination between rewarded and non-rewarded stimuli. To investigate the main hypothesis described above, I used a range of experimental approaches to interrogate the neuromodulation of the direct and indirect pathways by DA.

Chapter 2 describes the general methods, including methodology common to at least two experimental chapters.

Chapter 3 aims to elucidate the dissociable effects of D1R and D2R in the NAcS during reversal learning. I used a serial visual reversal-learning task and local administration of D1R and D2R antagonists while animals performed a serial visual reversal-learning task to assess the role of this region and receptors in different phases of reversal learning.

Chapter 4 aims to determine the role of D2R in reversal learning and their influence in modulating learning from positive or negative feedback. The recently developed valence-probe visual discrimination task (Alsiö et al., 2019) was used to dissociate different components of reversal learning and investigated the extent to which these were dependent on D2R systemically or locally in the NAcC and NAcS.

Chapter 5 aims to determine the synaptic location – pre- or post-synaptic – of D2R involved in the modulation of reversal learning. I used a post-synaptic probe compound (an A2AR antagonist) in a spatial probabilistic reversal-learning task in combination with D2R agents to test if the effects of D2R agonism in reversal learning were selectively mediated by striatopallidal D2R.

Chapter 6 aims to dissociate the role that RPEs generated in the nigrostriatal or mesolimbic pathways have in modulating reversal-learning performance. *In-vivo* optogenetics was used to stimulate the activity in both circuits during specific time points in a spatial probabilistic reversal-learning task to investigate the causal link between bursts of DA firing, behavioural performance, and the mediating role of these two pathways.

Chapter 2

General methods

This chapter describes methodology common to more than one experimental chapter. All other methods specific to individual experiments or further specifications to the methods given in this chapter are provided in the relevant chapter.

2.1 Subjects

Subjects were male Lister Hooded rats, housed in groups of four under temperature- and humidity-controlled conditions and a 12:12h dark cycle. Rats were allowed a minimum of 7 days of acclimatization to the animal facility before any procedure began. All rats were $\sim$300 g at the beginning of training and were maintained at 90% of their free-feeding weight by food restriction (19 g/day of Purina chow). Water was provided *ad libitum* in the home cage.

In Cambridge, all experimental procedures were subject to regulation by the United Kingdom Home Office (project License 70/7548) according to the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB).

At Boehringer Ingelheim, Germany, all experiments were performed in accordance with EU and German animal experiment legislation, under the license VVH-18-040-G granted by

local authorities, and under the direct supervision of Boehringer Ingelheim internal animal welfare officer.

2.2 Apparatus

2.2.1 Touchscreen operant chambers

The behavioural apparatus consisted of 28 touchscreen operant chambers (modified from Med Associates, Georgia, VT, USA), each measuring 29 $\times$ 31 $\times$ 24 cm with Plexiglas ceiling, front door and back panel. The floor was stainless steel bars separated by one cm from each other with a tray underneath. Access was through a hinged sidewall, secured with a latch during testing. Each chamber was equipped with a fan, house light (3 W), pellet dispenser, and magazine with light and photocell nose poke detector. The opposite wall was replaced with an infrared touchscreen monitor (29 $\times$ 23 cm).

2.2.2 Lever pressing chambers

The behavioural apparatus were eight operant chambers (Med Associates, Georgia, VT, USA), each enclosed within a sound-attenuating wooden box fitted with a fan for ventilation. Each chamber measured 31.4 $\times$ 25.4 $\times$ 26.7 cm with a Plexiglas ceiling, front door, and back panel. On one side of the chamber, a food magazine was centrally located and equipped with a light and photocell nose-poke detector. A pellet dispenser was connected to the magazine to deliver the reward: 45 mg sucrose pellets (5TUL, TestDiet, USA). Two retractable levers and a cue light above each lever flanked the magazine. On the same wall, a house light (3 W) was located to illuminate the chamber. The floor was made of stainless bars, each separated by one cm with a tray underneath. Access was through a hinged sidewall, secured with a latch during testing. The ceiling had a centered 5 cm diameter hole to allow electrical cables into the chamber.

2.3 Behavioural procedures

This section is divided between the visual reversal learning tasks, tested using touchscreen operant chambers at the University of Cambridge, and the spatial probabilistic reversal learning tested using lever-based operant chambers at Boehringer Ingelheim, Germany. In all the experiments, animals were tested once each day, 5-7 days a week, until completion of the study. Behavioural training started at least 2 days after food restriction. Before being exposed to the behavioural test chambers, rats received $\sim$30 sugar pellets in their home cage.

2.3.1 Visual reversal learning

Visual reversal learning was tested in touchscreen operant chambers. Performance required pre-training stages, training in visual discrimination, training in reversal learning and, finally, testing in the reversal-learning task (Fig. 2.1). Two different visual tasks were used: 1) serial reversal learning and 2) valence-probe visual discrimination (VPVD) task. Both tasks share pre-training, visual discrimination training, and the majority of the reversal-learning training stage. For this reason, the description of these tasks is common for both until the final stages of training and testing.

Training stages

Pre-training

The purpose of the pre-training sessions was to ensure that animals touched a start box at the bottom of the touchscreen, which in later stages would initiate the trial. Completion of this phase required approximately 2 weeks (Fig. 2.1A).

Rats were initially trained to touch the screens with daily sessions of 60 min or 100 trials. Pre-training consisted of five stages with gradually increased difficulty (Fig. 2.1B). Briefly, in stage 1, a large white horizontal square 'start-box' (15 $\times$ 9 cm) was presented in the bottom centre of the screen, and touching it was associated with reward (45 mg sucrose pellet; TestDiet 5UTL; Sandown Scientific, Middlesex, UK). The size of the 'start box' decreased throughout the stages until measuring 3 $\times$ 4 cm in stage 3. In stages 1-3, animals were moved to the next stage when reaching 100 responses/rewards per session.

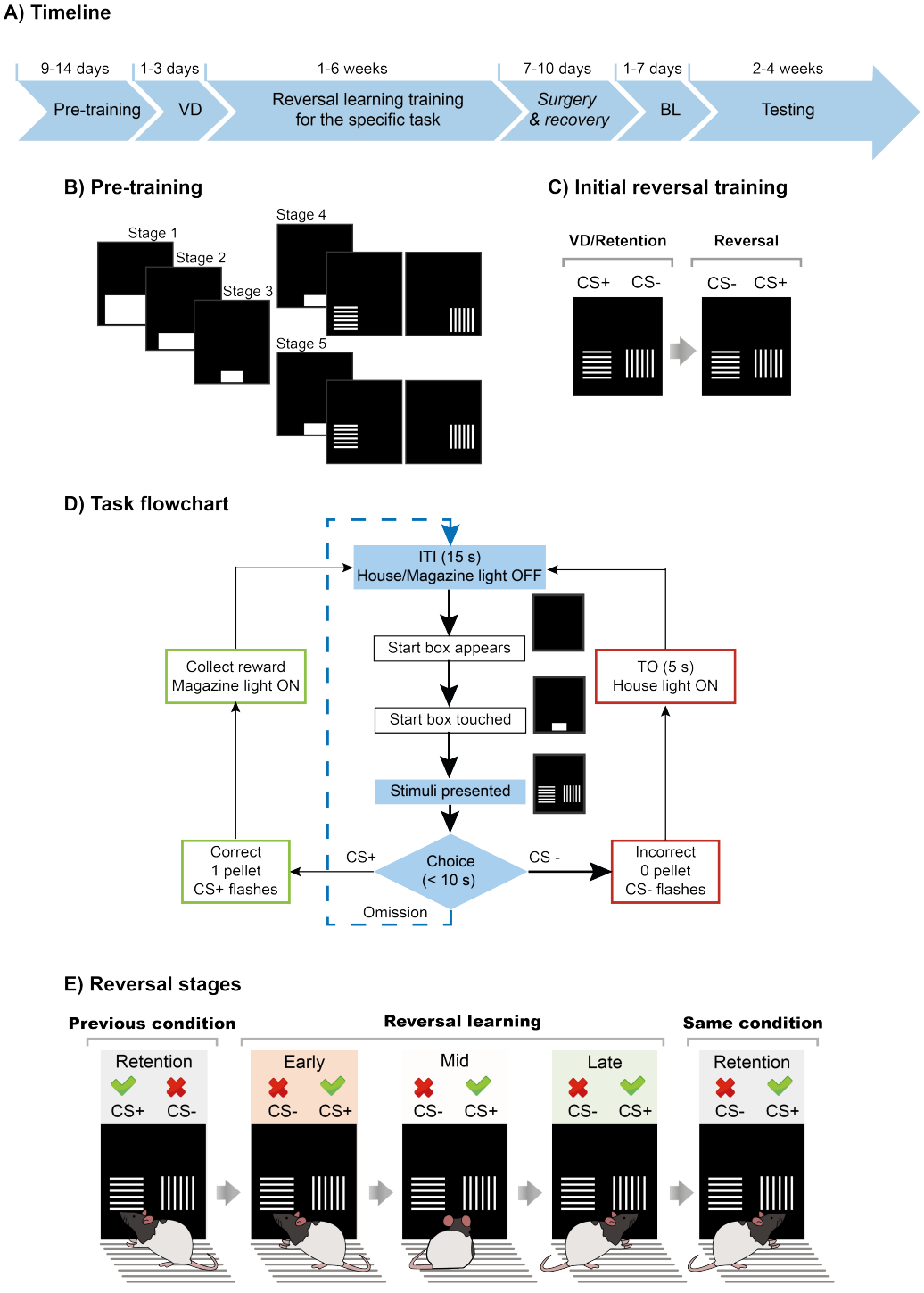


Fig. 2.1 Figure caption on following page.

Fig. 2.1 Overview of the touchscreen reversal-learning paradigm. A) Experimental timeline of the behavioural procedure, including pre-training stages, visual discrimination (VD), training in the relevant version of the reversal-learning task, surgeries and recovery when required by the experiment, baseline (BL) and behavioural testing with the relevant manipulation. B) Pre-raining stages. In stages 1-3 the white box decreases in size with stage level and becomes the start box. In stages 4 and 5 either the horizontal or the vertical stimulus is presented following touching the start box from stage 3. Stimuli are placed pseudo-randomly left/right. C) Representation of the stimuli presented during VD and initial, shared stages of reversal learning training. D) Flowchart of the visual discrimination and reversal task. E) Representation of a rat performing in the operant chambers.

In stage 4, touching the white box was not reinforced but led to the presentation of a visual stimulus (vertical or horizontal bars) with a pseudo-random spatial location, presented on the left or right of the touchscreen. The same stimulus was not displayed on the same side for more than three consecutive trials to avoid side-bias. Responding to the stimulus was reinforced, whereas the blank side led to the illumination of the house-light for a 5 s time-out (TO) period. After collecting the reward, an inter-trial interval (ITI) of 5 s was imposed. During stage 5, stimuli were presented slightly higher than in previous stages to avoid accidental touches e.g. with the tail. The criterion to move from stages 4 and 5 was $\geq$ 80% of correct responses per session.

Visual discrimination training

Following pre-training stages, subjects were trained on a visual two-choice discrimination task (Fig. 2.1D). In this version of the task, touching the square 'start box' triggered the simultaneous presentation of two stimuli (i.e. vertical and horizontal bars), placed pseudo-randomly on either the left or right side of the screen (Alsiö et al., 2015). The 'start-box' was used to ensure a central position of the animal before the choice phase. Responses to one stimulus (i.e. CS+) were associated with reward and collecting the reward initiated the next ITI. In contrast, responses to the other stimulus (i.e. CS-) were not rewarded and led to a house light-signalled TO. The response window after stimulus presentation was set to 10 s, after which the trial was deemed an omission and led to a new ITI. The session ended after 250 trials, 50 rewards or 1 h, whichever came first.

Criterion for discrimination learning was set at 24 correct responses out of 30 consecutive trials. Once acquired within any session, rats were given a retention session with the same reward contingencies to ensure they had reliably acquired the visual discrimination.

Reversal learning training

Following acquisition of visual discrimination, animals were trained on the serial visual reversal learning task (Fig. 2.1C, E)). After the discrimination and retention sessions, contingencies reversed so the previous CS+ was then CS- and *vice versa*. Rats were required to respond to the new CS+ until reaching the discrimination criterion ($\geq$ 24/30 correct responses). After reaching criterion, an extra retention session was run.

Following reversal training, animals were tested in either the serial reversal-learning task (see section 1.3.2) or the valence-probe visual discrimination (VPVD) reversal task (see section 1.3.3). For the serial reversal-learning task, but not the VPVD task, additional reversals were given until rats were able to attain the discrimination criterion within three daily sessions. A retention session was run before each reversal and after reaching the criterion (Fig. 2.1D), both in training and testing.

Serial reversal learning task

When animals met the criterion of completing a reversal within 3 sessions, they underwent surgery for cannula implantation. Following recovery from surgery, and prior to testing, rats were tested with a single reversal learning session as a baseline to ensure stable serial reversal performance. The following session was a retention session, followed the following day by a reversal session. Stimulus-outcome conditions were reversed prior to the reversal session, so that the previous rewarded stimulus was now non-rewarded and *vice versa*. Animals were tested on multiple sessions until reaching the discrimination criterion (> 24/30 correct trials) within one session. A retention session was ran the following day. This process was repeated for each drug and dose of each Latin-square (LSQ) or crossover design. In serial reversal learning, the effects of each dose of drug were tested using a within-subject's design. Each reversal, including retention sessions, lasted seven days or less.

VPVD task

Studies in patients with neurological disorders have established alterations in adapting behaviour on the basis of positive and negative feedback, including subjects with depression or with Parkinson's disease off-medicine (Elliott et al., 1997; Frank et al., 2004). The typical

touchscreen visual discrimination procedures require the subjects to discriminate between CS+ and CS- (Brigman et al., 2013; Horner et al., 2013). The recently developed VPVD task (Alsiö et al., 2019) is based on two-choice visual discrimination and reversal learning, but extends the traditional framework by introducing a third probe neutral stimulus ($C_{50/50}$) that is rewarded 50% of the times (Fig. 2.2). By examining animals' performance on these 'probe trials', it is possible to determine the impact of positive and negative association on task performance.

A)

	VD	Reversal	Learning strategy	Trials/session
Standard	A+ > B-	B+ > A-	Approach/Avoid	150
Positive probe	A+ > $C_{50/50}$	B+ > $C_{50/50}$	Approach positive	25
Negative probe	$C_{50/50}$ > B-	$C_{50/50}$ > A-	Avoid negative	25

B)

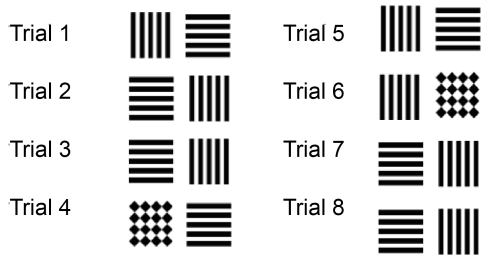


Fig. 2.2 Overview of the trial types and trial structure in the VPVD reversal task. A) Regular two-choice trials during both visual discrimination (VD; A+ > B-) and subsequent reversal learning (A- < B+) are alternated with probe trials. During such trials, a third neutral stimulus that is rewarded 50% of the times ($C_{50/50}$) is presented against the positive or the negative stimulus. B) Illustration of a possible set of trials presented during the session. The order and combination of stimuli are pseudo-randomised, including probe trials, although these are never presented during the first trial.

Following reversal learning training, standard discrimination trials (CS+ *vs* CS-) were interleaved with the third probe neutral stimulus ($C_{50/50}$, "Diamonds"). That is, $C_{50/50}$ was optimal over CS- but not over CS+ (Fig. 2.2A). Each probe trial was randomly presented once every eight trials (2/8 probe trials in total), although never on the first trial (Fig. 2.2B). The remaining six trials were standard discrimination to ensure animals' learning was mainly driven by the deterministic conditions. No omissions were allowed in this task in order to ensure rats had completed the probe trials. Instead, a tone was presented every time a trial

was rewarded. Rats trained for five sessions on the same CS+ and CS- during the training reversal above (i.e. "Horizontal" *vs* "Vertical") in combination with the $C_{50/50}$ stimulus.

After completing the discrimination, they were trained on the same conditions but with a novel pair of stimuli ("Forward slash" *vs* "Backslash", counterbalanced across rats). CS+ and CS- were referred to as A+ and B- during visual discrimination. Training continued for a minimum of 5 sessions and was extended for those animals that did not reach 80% correct on the standard trials (A+ > B-). Next, rats ran a retention session preceded by vehicle infusions (baseline). On the following day, rats were matched for stimulus-reward contingencies, performance on the baseline day and during reversal training and were pseudo-randomly allocated to a drug group. The stimulus-reward contingencies were reversed for the deterministic stimuli (A- < B+) at the beginning of the session and the reversal phase continued for 10 days. The $C_{50/50}$ stimulus value remained the same. A summary of the task is shown in Table 2.1.

2.3.2 Spatial probabilistic reversal learning

Training stages

Animals were first habituated to the chambers for 15 min with three sugar pellets placed in the magazine before the start of the session. In all stages (Table 2.2), trials began with the illumination of the magazine light. Animals were then trained in stage 01 or 'conditioning', which consisted of 60 min sessions of 40 trials to learn that pressing the lever delivered a reward pellet. Both levers were presented simultaneously, and if one of them was pressed, three pellets were delivered; if not, only one pellet was delivered. The criterion to move to the next stage was the completion of 40 trials. All the following training stages consisted of sessions of 60 min or 120 trials, whichever came first, and incorporated an inter-trial interval (ITI) of 10 sec, a time-out (TO) of 5 sec after an incorrect response, and a limit hold (LH) of 30 sec, after which levers were retracted and the trial was considered as an omission. Criterion to move to the next stages was the attainment of $\geq$ 80 correct trials. In the next stage, stage02 or 'must press', animals were trained to press the lever to receive a reward by receiving one pellet when one of the two levers was touched, or none if not. After 60 trials, the preferred lever was retracted to force animals to press the adjacent lever and so not develop a side bias. The following stage, stage03 or 'must initiate', had the same set-up as

Summary VPVD

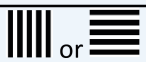
Stage	Stimuli	Duration (in weeks)	Purpose
Pre-training (stages 1-5)		2	Learning to touch the screen and initiate the trial
Training visual discrimination		1	Learning to make a decision between two stimuli to be rewarded
Training first reversal learning		2	Learning how to respond when contingencies are reversed and gain stability in the performance
Training first visual discrimination in the VPVD task		1 (5 sessions)	Introduction to the VPVD task to learn to discriminate a new third neutral stimulus
		Surgeries	
Acquiring second visual discrimination in the VPVD task		1 (5 sessions)	Learning the conditions with the set of stimuli that will be used during testing
Testing reversal learning		2-3 (10 sessions)	Testing behavioural flexibility

Table 2.1 Summary of the VPVD task stages, including stimuli presented, expected duration and purpose of the stage. When required, surgeries are performed in between achieving criterion for the first discrimination in the VPVD task and the beginning of the second discrimination. Training is achieved with "Horizontal"/"Vertical" bars as A+ and B-, which are rewarded 100% and 0% of the times. To train in the VPVD task, a new third stimulus ("Diamonds") is introduced, which is rewarded 50% of the times. For testing, a new set of A+/B- stimuli is introduced to avoid behavioural artefacts due to previous experiences with the stimuli. The animals are trained in a second discrimination using "Forward Slash"/"Backslash", which follow the same principal as the previous vertical and horizontal bars. Diamonds stimulus remains unaltered

the previous one but animals had to nose poke into an empty magazine to initiate the trial. A light in the food magazine was flashed to indicate the difference from pellet delivery, which was indicated with a steady light. The last training stage was stage04, a one-sided version of the full task. The aim was to avoid side-bias and that both levers could be rewarded in a probability basis. For this, both cue lights would turn on and after 3 sec, only one of the levers came out, which was rewarded 80% of the times. After 30 consecutive trials, the levers

shifted. At least two sessions reaching criterion ($\geq$ 80 correct trials) were required to move to probabilistic reversal learning (PRL) task training.

		Stage00 Habituation	**Stage01** Conditioning	**Stage02** Must press	**Stage03** Must initiate	**Stage04** On-sided full task	**Stagev01** Full task
Aim		Get habituated to the chambers	Get familiar with the levers and that pressing them is rewarded	Learn that levers must be pressed to receive a reward	Learn to initiate the trial by nosepoking into an empty magazine	Avoid side-bias and learn that both levers can be rewarded in a probabilistic manner	Training on the final version of the task until reaching stable performance
Procedure		Five pellets are placed in the magazine. The beginning of the trial is signalled by the house light flashing, the magazine light turning on, and delivery of 2 free pellets.	Both stimuli lights turn on and the levers come out. If the levers are touched, 3 pellets are delivered. If not, 1 pellet.	Both stimuli lights turn on and the levers come out. If the levers are touched, 1 pellet is delivered. If not, 0 pellets. After 60 trials the preferred lever is skipped.	Same as stage02, but after ITI, the magazine light flashes to indicate the need of nosepoking to continue the task.	Same as stage03, but after ITI, stimuli lights turn on, indicating that after 3 s one of the levers comes out (80% rewarded). Same lever for 30 trials, and then they shift. NO levers are skipped after 60 trials.	Same as stage 03, bit both levers come out after 3 s of lights on. One lever is optimal (i.e. 805 rewarded) and the other is sub-optimal (i.e. 20% rewarded). After 8 consecutive responses on the optimal lever, contingencies reverse.
Timings	Length	15‘	60’	60’	60’	60	60
	ITI	NA	10”	10”	10”	10	10
	TO	NA	5”	5”	5”	5	5
	LH	NA	30”	30”	10”	10	10
Criterion		1 session	Complete 40 trials	Complete 120 trials and ≥ 80 rewards	Complete 120 trials and ≥ 80 rewards	Complete 120 trials and ≥80 rewards	≥ 2-3 rewards over 2-3 sessions
Expected sessions		1	1-2	1-4	1-4	2-4	10

Table 2.2 Overview of the training stages in the PRL task, including habituation (Stage00), the four learning stages (Stage01-Stage04), and the full version of the task (StageV01).

Probabilistic reversal learning

PRL procedures, although are recent, have been widely used to study flexible behaviour in rodents and humans (Bari et al., 2010; den Ouden et al., 2013; Ineichen et al., 2012). PRL paradigms often involve the release of probabilistic feedback, so that the correct choice is spuriously punished, and the incorrect choice is spuriously reinforced (Fig. 2.3B).

The procedures used in the present study (Fig. 2.3) were modified from those described in (Bari et al., 2010) for the use of retractable levers instead of nose poking holes. Daily sessions consisted of 200 trials or 60 min, whichever came first, including ITI of 10 sec, TO of 5 sec and LH of 10 sec.

At the start of each session, one of the two levers was randomly selected to be optimal or suboptimal. Sessions began with two free pellets in the magazine and illumination of the

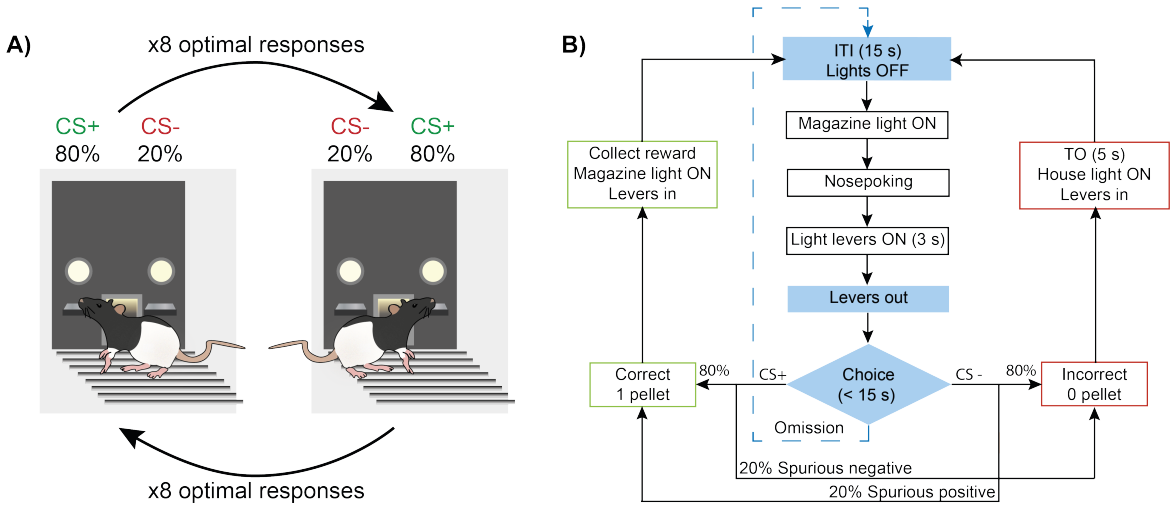


Fig. 2.3 Overview of the PRL task. A) Illustration of the rats performing in the lever-pressing operant chambers. After eight consecutive correct trials on the optimal lever (CS+; 80% rewarded), conditions reverse, so that the previously optimal lever is now suboptimal (CS-; 20% rewarded), and *vice versa*. B) Flowchart of the PRL task.

magazine light. After nose poking, the magazine light was extinguished and the two cue lights turned on indicating that three seconds later the levers would be available. A response to the optimal lever delivered a single reward pellet on 80% of the trials, whereas a response to the suboptimal lever gave reward on only 20% of the trials. A failure to press a lever within the LH led to their retraction and termination of the trial, which was noted as an omission. After eight consecutive correct trials (i.e. pressing the optimal lever regardless of it being reinforced or not), the contingencies were reversed, so that the previous optimal lever was now suboptimal and *vice versa* (Fig. 2.3A). This pattern was repeated over the session. Animals were trained until they could achieve at least three reversals per session over three consecutive sessions. Once this criterion was met, rats underwent testing.

2.4 Surgeries

Surgical procedures were performed following standard stereotaxic techniques for cannula implantation (Chapters 3 and 4) and expression of viral vectors (Chapter 6). Methods for viral vector infusion studies are presented in Chapter 6.

Rats were anaesthetised using isoflurane in 5% oxygen and secured in a stereotaxic frame fitted with atraumatic ear bars. Anesthesia was maintained at 2.5% isoflurane. Baytril (1 mg/kg; 100 mg/ml; Bayer, Germany) and metacam (1 mg/kg; 5 mg/ml; Boehringer Ingelheim, Germany) diluted in distilled water 1:1 were injected subcutaneously prior to surgery.

2.4.1 Cannula implantation

Specifications of guided cannulas and coordinates are included in each chapter.

Flat skull was ensured by measuring DV from lambda and Bregma. Coordinates were according to Paxinos and Watson (1998). Four metal screws and dental cement secured the guide cannulas to the skull. Obturators were introduced in the guide cannulas and protected with a dust cap. After surgery, animals were single-housed, received metacam orally (1.5 mg/ml at 1 mg/kg) during three days post-surgery, and allowed $\geq$ 7 days of recovery.

2.5 Drug microinfusions

After a baseline session, animals received a mock infusion, consisting of introducing the injectors to the guide cannula without infusion of liquid. The injectors were connected to the lines with filtered saline to prevent air entrance to the brain. The day prior to testing, they received a retention session preceded by an infusion of vehicle to allow habituation to the infusion procedure (vehicle composition described in each chapter including infusions i.e. Chapters 3 and 4). During infusions, rats were placed on the lap of the researcher and were gently restrained. Animals were infused bilaterally with a total volume of 0.5 μl/side over 2 min. Injectors (Plastics One, 28-gauge, USA) were left in place 1 min before and after the infusions, and rats were placed in their home cages before putting them in the operant chambers to test the behavioural effects of each drug. Timings and injectors' extension below the guide cannula are specified in each experimental chapter. Doses were pseudo-randomised depending on baseline performance on the day prior to testing.

2.6 Histologies for cannula tip placement

Following completion of the behavioural procedures, animals were anaesthetised with a lethal dose of pentobarbitone (∼2 ml; Boehringer Ingelheim, Germany) and perfused transcardially with 0.01 M phosphate buffer saline (PBS) followed by 4% paraformaldehyde (PFA). Brains were removed, post-fixed in PFA for 24 h and dehydrated in 30% sucrose in 0.01 M PBS. Brains were sectioned coronally at 60 μm. They were then mounted, stained with cresyl violet and cover slipped to verify injector-tip placements within inside the NAcC or the NAcS. Only animals with correct cannula placement were included in the analysis.

2.7 Statistical methods

Statistical tests were performed with R, version 1.2.1335 (RStudio Inc). The main dependent variables analysed are specified in each chapter. Normality was confirmed with a quantile-quantile plot (Q-Q plot). For repeated measures variables, sphericity was checked with the Mauchly's test for sphericity and corrected with Greenhouse-Geisser when required.

Data were then subjected to ANOVA or Linear Mixed-Effects Model analysis using the lmer package in R. When significant interactions were found, analysis was followed by post-hoc pairwise comparisons using the emmeans package in R. If any number of comparisons were made among the set of groups, a Tukey's procedure was used to correct p-values. Instead, if treatment groups were compared against a control group, Dunnett's test was applied to correct p-values. Experiments in the VPVD task considered dose as a between-subject effect. In the PRL task, reversible treatments (e.g. drug doses and light conditions in optogenetics) were analysed as within-subjects, whereas irreversible manipulations (e.g. opsin-expressing groups) were analysed as between-subjects factors. Significance was considered at $\alpha = 0.05$.

Chapter 3

Accumbal dopamine D1 and D2 receptors differentially modulate distinct phases of serial visual reversal learning

3.1 Introduction

Converging evidence from reversal learning tests implicates DA as an important modulator of such flexible behaviour. For instance, systemic blockade or agonism of D2R impairs reversal learning in vervet monkeys and rats (Boulougouris et al., 2009; Lee et al., 2007), while D2R knockout mice show deficiencies in initial visual discrimination and in reversal learning (Kruzich and Grandy, 2004). In contrast, pharmacological activation of D1R impaired early phases of reversal learning (Izquierdo et al., 2006), whereas D1R antagonism did not alter reversal learning performance (Lee et al., 2007). In healthy humans, repeated variations in the DA transporter gene, *DAT1*, have been linked to performance during the early, perseverative phase of reversal learning, when prior beliefs about the stimulus-reward outcomes still guide behaviour, whereas accuracy during later phases, when new learning takes place, showed no such link (den Ouden et al., 2013).

The main subregions of the dorsal striatum, namely the caudate nucleus and the putamen in primates and the DMS and DLS in rodents, have also been differentially linked to reversal learning. Recent evidence suggests that pharmacological inactivation of the putamen and

caudate nucleus differentially affect serial visual reversal learning in marmoset monkeys (Jackson et al., 2019). D2R availability in these subregions of vervet monkeys is associated with reversal learning performance (Groman et al., 2011). Importantly, previous research of ours, presented in Dr Leanne Young's thesis (Young, 2019), showed a complimentary role of the DMS and DLS depending on the phase of reversal learning: early, mid or late (Sala-Bayo et al., 2020). D2R antagonism into the DMS increased errors to reach criterion to complete reversal in the mid phase, whereas D2R antagonism into the DLS increased overall errors, especially in the early and late phases. In contrast, D1R antagonism in these regions did not affect reversal-learning performance. However, less is known about the role of these two receptors in the ventral striatum.

Previous studies have shown that increased dopaminergic tone in the NAc, or infusions of a D2R agonist (quinpirole) into this area impaired reversal learning in rats, whereas infusions of a D1R agonist (SKF81297) disrupted set-shifting by increasing perseverative behaviour (Haluk and Floresco, 2009). Lesions of the NAc disrupted initial stimulus discrimination and reversal learning (Annett et al., 1989; Taghzouti et al., 1985), including spatial, but not visual, reversal learning in monkeys (Stern and Passingham, 1995), and pharmacological inactivation impaired probabilistic learning in rats (Dalton et al., 2014). However, other studies report no effect of NAc interventions on such flexibility (Castañé et al., 2010; Schoenbaum and Setlow, 2003).

This discrepancy may be explained by the heterogeneity of the NAc with the NAcC and NAcS often being suggested to play opposite roles in modulating behaviour, and contributing differentially to e.g. attention (Corbit et al., 2001; Floresco et al., 2006) and impulsivity-related behaviours (Besson et al., 2010; Economidou et al., 2012; Sesia et al., 2008). For instance, inactivation of the NAcS impaired probabilistic reversal performance in rats, identifying a key role for this nucleus in using probabilistic reward feedback to facilitate discriminative learning and flexibility, whereas inactivation of the NAcC, while not affecting performance accuracy did cause a general slowing of approach toward the response levers (Dalton et al., 2014). Previous results from the lab (Sala-Bayo et al., 2020) reported that infusions of a D2R antagonism (raclopride) into the NAcC selectively improved performance in reversal learning by reducing the number of errors to reach criterion in the early, perseverative phase. In contrast, no effect on reversals was observed when blocking D1R within this region.

Taken together, this evidence suggests that elevated DA activity in the ventral striatum leads to impaired reversal learning. However, the role of D1R and D2R in the NAcS in visual reversal learning or of their involvement in its different learning phases is unknown.

3.2 Aims, approaches, and hypotheses

The aim of this study was to investigate whether D1R and D2R in the NAcS differentially affect reversal learning, and on the different phases of reversal learning. To achieve this goal and better understand such dissociations, I explored the behavioural effects of local administration of a D2R antagonist and a D1R antagonist using a serial visual reversal-learning task on touchscreen operant chambers. It was hypothesised that the role of DA receptors in the NAcS would be different, perhaps complementary to the improvement in reversal performance observed in the early phase when antagonising D2R in the NAcC.

3.3 Methods

All experiments were performed at the Department of Psychology, at the University of Cambridge, UK.

3.3.1 Subjects

Animals were kept under the conditions specified in Chapter 2 (see section 2.1). A total of 22 male Lister-Hooded rats (Charles River, UK) were used for this study. Rats were housed in groups of four, but single-housed following implantation of guide cannulas. The number of animals used for each drug and region is shown in Fig. 3.1B.

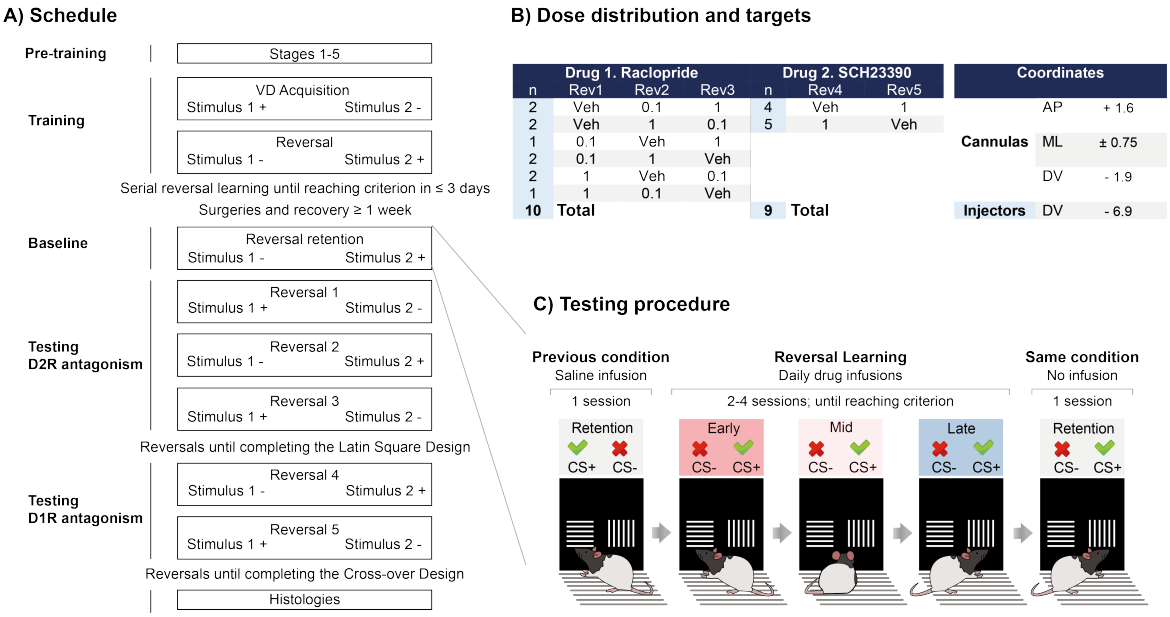


Fig. 3.1 Experimental design. A) Behavioural training and testing protocol. The rewarded stimulus is represented as a + and the unrewarded stimulus as a -. Stimuli were vertical or horizontal bars and were counterbalanced as CS+ or CS- across rats. B) Dose distribution across reversals (Rev) and targeted coordinates for the NAcS. The n column includes subgroup and final group size for the different DA receptors antagonists: raclopride (D2R) and SCH23390 (D1R). The same rats received both drugs, although one had to be excluded before administration of SCH23390. Doses are presented as μl/μl. AP and ML were measured from bregma and DV from dura. C) Flowchart of the testing procedure and phases of reversal learning. Phases depended on performance within sessions. After reversal, during the early phase performance was lower than 11 correct trials out of a set of 30 trials, as animals tended to perseverate on the previously CS+, now CS-. After some trials, performance improved, and animals reached the so-called mid, intermediate or random phase, before reaching the late or learning phase, in which they have learnt to approach the new CS+ (> 19/30 correct responses).

3.3.2 Behavioural procedures

Pre-training

Rats were initially trained to respond to touchscreens operant chambers. See Chapter 2 (section 2.3.1) for details on pre-training stages.

Visual discrimination training

After the initial training stages, subjects were trained on a visual two-choice discrimination task. See Chapter 2 (section 2.3.1) for details on visual discrimination training.

3.3.3 Serial visual reversal learning

Following acquisition of visual discrimination, animals were trained in serial visual reversal learning. See Chapter 2 (section 2.3.1).

3.3.4 Surgeries

Bilateral 22-gauge guide cannulas (PlasticsOne, Sevenoaks, UK) were implanted in the NAcS (n = 22). following standard stereotactic techniques. For more details see Chapter 2 (section 2.4). Coordinates for NAcS guides were AP +1.6, DV -1.9 and ML $\pm$ 0.75 (Fig. 3.1B).

3.3.5 Drugs

The DA D2R antagonist raclopride (Tocris Bioscience, Bristol, UK) and the D1R antagonist SCH23390 (Sigma-Aldrich, Dorset, UK) were dissolved in physiological saline. The drugs were infused into the NAcS at the doses of 0, 0.1 and 1 μg/μl of raclopride and 0 and 1 μg/μl of SCH23390. Aliquots were frozen at -80°C in the quantities required for each testing day.

3.3.6 Microinfusions

Intracerebral infusions took place according to the procedure described in Chapter 2 (see section 2.5). Doses of raclopride were administered in a LSQ design and of SCH23390 in a cross-over design. Although all doses order for each rat were randomized and controlled for baseline performance, all animals were infused with raclopride first and received SCH23390 infusions only after the raclopride LSQ was completed (i.e. whereas the individual doses

within each experiment was counter-balanced, we did not counter-balance between raclopride and SCH23390). On average, each animal received 4 sessions with infusions per reversal cycle (vehicle on retention and three days of drug dose or vehicle during reversal), thus around 12 infusions in the raclopride LSQ, and eight infusions in the cross-over study for SCH23390.

Injectors from PlasticOne (28-gauge, USA) extended 5 mm below the guide (final DV – 6.9). The injectors were left in place for 1 min both prior to and after the infusion. Rats were returned to their cage for 5 min before the start of the session.

3.3.7 Histologies

See Chapter 2 (section 2.6) for details on histologies.

3.3.8 Data analysis

See Chapter 2 (section 2.7) for details on statistical analysis. In this chapter, the main dependent variables were the number of errors and trials to criterion ($\geq$ 24/30 correct responses). Omissions, latencies to respond and latencies to collect the reward were additionally analysed. Data from each reversal were collapsed over days. Trial outcomes were classified in three different phases: early, mid or late, depending on the performance over a running window of 30 consecutive trials. If animals had a significant bias (binomial distribution probabilities) towards the previously positive stimulus (< 11/30 correct responses), performance was considered to belong to the early phase, in which animals exhibited mainly perseverative responses. If their performance instead showed a significant preference for the currently rewarded stimulus (> 19/30 correct responses) it was considered as the late phase, in which animals moved closer to criterion for learning the reversed contingency. Performance in-between these thresholds was classified as intermediate or mid-phase, prior to acquisition of the new learned association. Data from all trials after the rats had reached the final learning criterion ($\geq$ 24/30 correct responses) were excluded from the analysis.

To ensure normality, errors were square-rooted and latencies log-transformed. Data were then subjected to Linear Mixed-Effects Model analysis with the lmer package in R. For each region, the model contained two fixed factors (dose, phase) and one factor (subject)

modeled as a random slope to account for individual differences between rats across phases (i.e. individual learning curves). When significant two-way dose × phase interactions were found, analysis was followed by *post-hoc* Dunnett's corrected pairwise comparisons.

3.4 Results

3.4.1 Histology

The ventral-most locations of injectors are included in Fig. 3.2. Rats were excluded from the study if the injector cannulas were positioned outside the target areas (n = 3). Final group sizes with verified injector positions for each of the drug groups and targeted coordinates are shown in Fig. 3.1B.

Effects of intra-striatal infusions of the D2R antagonist raclopride and the D1R antagonist SCH23390

Analysis for both raclopride and SCH23390 treatments substantiated that the effect of drugs varied across phases of the reversal task. Fig. 3.2 shows that whereas local infusions of the D1R antagonist SCH23390 improved early to mid stages in reversal learning, no effects were observed with the D2R antagonist raclopride into the NAcS.

Analysis on the number of errors committed after SCH23390 infusions identified a significant dose × phase interaction after NAcS infusions ($F_{2,\ 31.997} = 25.516$, $p < 0.001$). *Post-hoc* analyses showed that D1R antagonism into the NAcS selectively decreased perseveration in the early phase compared with the vehicle condition ($p < 0.001$; Fig. 3.2C), and followed a trend in the mid phase ($p = 0.054$). In contrast, there was a dose × phase interaction for errors after local infusions of raclopride ($F_{4,63.005} = 3.813$, $p = 0.008$), but pairwise comparisons revealed that no dose differed from the vehicle-control group (Fig. 3.2B). There was thus no significant effect of raclopride when infused into the NAcS.

Effects on trials to criterion were similar to the ones from errors showed above. Infusion of SCH23390 decreased the total amount of trials to reach criterion. Mixed-Effects Model showed a significant dose × phase interaction ($F_{2,32} = 20.323$, $p < 0.001$). *Post-hoc* analysis

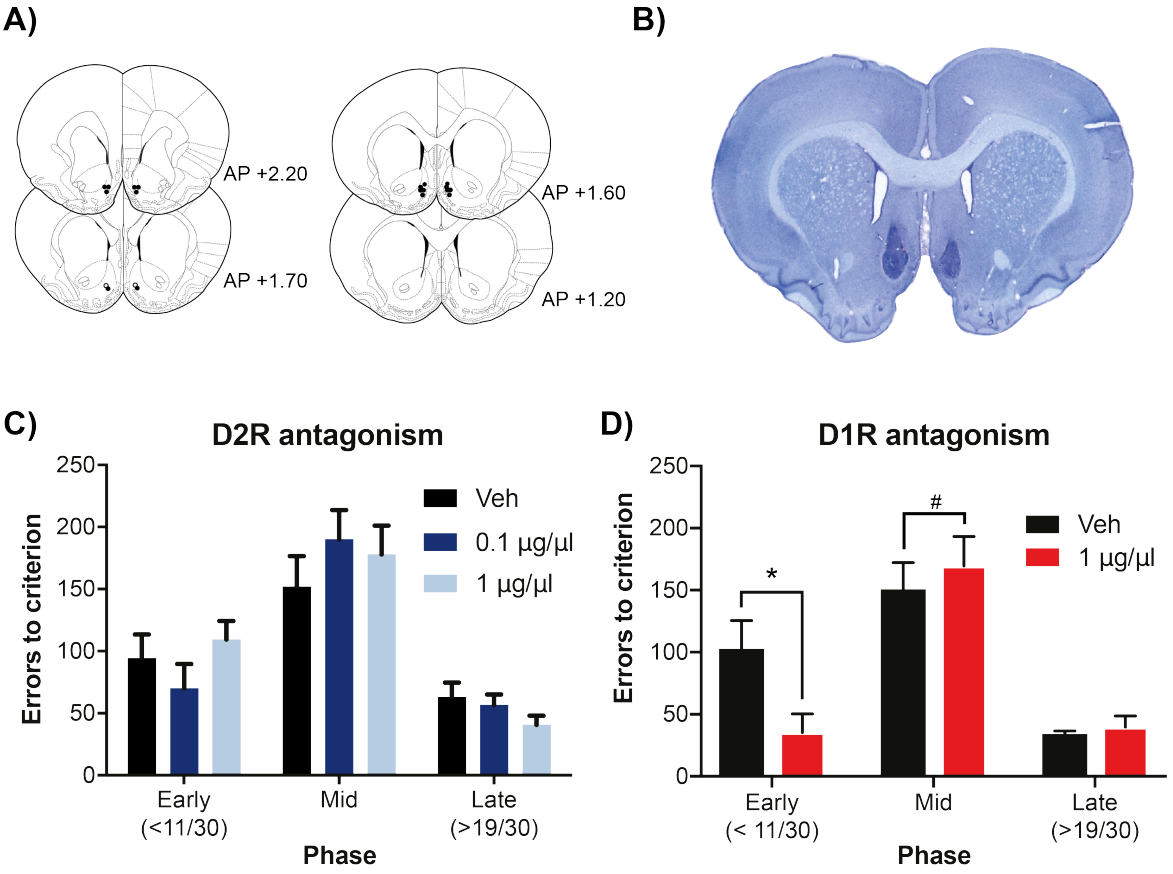


Fig. 3.2 D1R antagonism improves reversal-learning performance in the early phase. A) Injector tip placements. Closed circles represent rats that received both raclopride and SCH23390; open circles represent rats that received only raclopride. B) Example of bilateral injector tip placement. C) Errors to criterion by phase – early, mid and late – after the D2R antagonist, raclopride, in the NAcS. D) Errors to criterion by phase – early, mid and late – after the D1R antagonist, SCH23390, in the NAcS. Errors until reaching criterion of a high performance (> 24/30 correct responses) are collapsed over reversals. Data shown as mean $\pm$ SEM. $^{\#}$ p $\sim$0.05; * p < 0.05 *vs* vehicle treatment.

revealed this effect was due to the infusion of drug during the early phase ($p < 0.001$), tending to persist in the mid phase ($p = 0.086$). No effects were observed after infusion of raclopride.

Table 3.1 shows that in the NAcS, neither SCH23390 nor raclopride affected the number of omissions (dose $\times$ phase: $F_{2,\,35.007} = 2.112$, $p = 0.136$; $F_{4,\,72} = 0.830$, $p = 0.511$). SCH23390 infusions also prolonged the latencies to collect the reward and to respond to the stimuli regardless of the phase (dose in collect: $F_{1,\,31.062} = 99.382$, $p < 0.001$; dose in respond: $F_{1,\,31.082} = 7.838$, $p = 0.009$). Raclopride had no effect on these variables (Table 3.1).

Drug	Dose	Omissions			Latencies to respond			Latencies to collect		
		Early	Mid	Late	Early	Mid	Late	Early	Mid	Late
SCH	Veh	0.13 ± 0.13	0.50 ± 0.38	0.00 ± 0.00	997 ± 98	960 ± 99	974 ± 120	1283 ± 132	1166 ± 102	1055 ± 121
	1	0.00 ± 0.00	1.00 ± 0.42	0.88 ± 0.30	*1030 ± 83* [a]	*1053 ± 58* [a]	*1210 ± 58* [a]	*3086 ± 475* [b]	*2342 ± 336* [b]	*2194 ± 336* [b]
Raclopride	Veh	1.00 ± 0.70	0.50 ± 027	0.10 ± 0.10	970 ± 61	842 ± 68	816 ± 58	1889 ± 186	1524 ± 134	1253 ± 134
	0.1	0.30 ± 0.15	0.20 ± 0.22	0.10 ± 0.10	656 ± 105	809 ± 108	838 ± 90	1346 ± 190	1423 ± 134	1123 ± 138
	1	0.30 ± 0.15	0.50 ± 0.22	0.10 ± 0.10	972 ± 105	865 ± 108	853 ± 90	1575 ± 190	1396 ± 134	1166 ± 138

Table 3.1 Effects on omissions and latencies to respond or to collect the reward after D1R antagonist SCH23390 or D2R antagonist raclopride into the NAcS. Doses are presented as μg/μl; latencies in ms. Data are shown as mean ± SEM. [a] $p < 0.01$ *vs* vehicle treatment; [b] $p < 0.001$ *vs* vehicle treatment.

3.5 Discussion

This study demonstrates dissociable effects on visual serial reversal learning of D2R and D1R antagonists locally infused into the ventral striatum depending on different learning phases of the task (i.e. the early, perseverative phase *versus* new learning phases). An important overall finding was that whereas D1R antagonism into the NAcS improved early stages of reversal learning (Fig. 3.3 for visual interpretation), D2R antagonism failed to alter behaviour. This shows the existence of different roles of DA receptor signalling within the accumbal structure when stimulus-reward contingencies change.

This work was the continuation of my BSc project embedded in Dr Leanne Young's PhD (Young, 2019), in which we investigated the effect of the same DA receptors antagonists, raclopride and SCH23390, into the DMS, DLS and NAcC (Sala-Bayo et al., 2020). Briefly, and in opposition to the effect in the NAcS, we observed an impairment in the mid phase after D2R antagonism into the DMS, whereas it was observed across all phases, including early and late, when into the DLS. No effects were detected after D1R antagonism, suggesting a complementary role of these regions in modulating reversal learning via D2R. In the NAcC, blocking D2R improved performance during early phases of reversal learning, whereas D1R antagonism did not alter the amount of errors to reach criterion, in contrast to what we observed when targeting the NAcS.

These findings increase our understanding of the neural mechanisms modulating cognitive flexibility, and are in general consistent with previous data on humans with Parkinson's

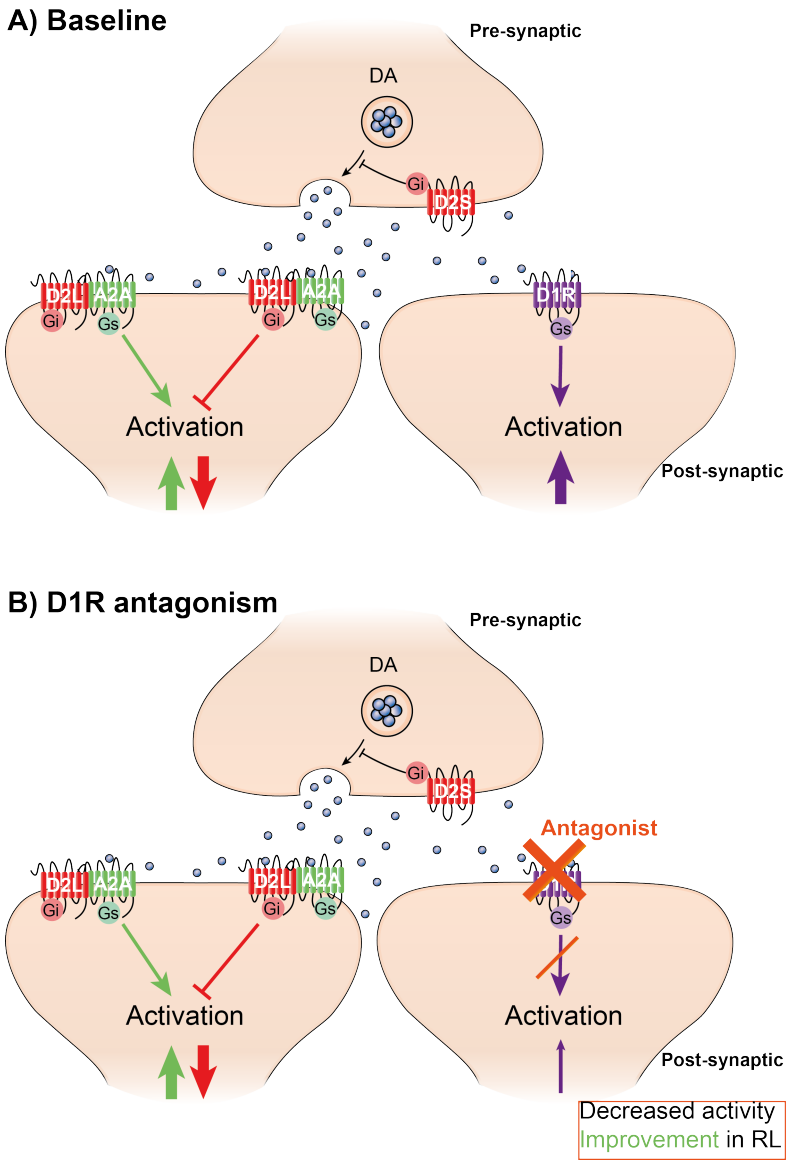


Fig. 3.3 Representation of the potential effects of D1R antagonism in NAcS MSNs. A) Baseline signalling by D1R- and D2R-expressing MSNs. DA is released by pre-synaptic neurons expressing D2S as auto-receptors. DA can join D1R and D2R (mainly D2L) in post-synaptic neurons and activate their signalling cascade. A2AR co-express with D2R in post-synaptic neurons. B) Interpretation of the effect of D1R antagonism by SCH23390. The antagonist may join D1R and block their downstream signalling, resulting in lower cell activation, and decreased pathway signalling, which improved reversal learning performance by enhancing learning in early perseverative stages.

disease (Cools et al., 2007; Dagher and Robbins, 2009) indicating that excess DA activity may often be detrimental for reversal performance in the NAc, whereas intact DA function

in the dorsal striatum is necessary for efficient reversal learning, as supported by data from non-human primates (Clarke et al., 2011; Groman et al., 2011).

Considering not only the NAcS results, but also the NAcC findings presented in Dr Young's thesis (Young, 2019) and in Sala-Bayo et al. (2020), the positive changes in reversal performance after local infusions relied on both the accumbal subregion and the subtype of DA receptor, and they were selective for the early phase of reversal learning. Whilst D1R antagonism in the NAcS reduced the number of perseverative errors, this improvement was only found after D2R antagonism in the NAcC. This double dissociation refines previous studies revealing for example that elevated dopaminergic levels in the NAc impair reversal learning (Verharen et al., 2018), and that D2R agonism in the NAc is detrimental for behavioural flexibility (Haluk and Floresco, 2009; Verharen et al., 2019). This might be relevant for the DA overdose hypothesis of pathologic cognitive impairments associated with dopaminergic drug treatment in Parkinson's disease (Swainson et al., 2000), since our data indicate that such effects are modulated by D1R in the NAcS and D2R in the NAcC. Nonetheless, since the antagonists employed here only block endogenous DA, our data also indicate that DA signalling at D1R in the NAcS and D2R in the NAcC supports perseverative responding in visual reversal learning, maybe by inadequately maintaining the previous stimulus-reward association (Flagel et al., 2011) or Pavlovian conditioned approach (Fraser et al., 2016). Inactivation of the NAcS can also improve numerous forms of behavioural flexibility, including spatial reversal learning (Aquili, 2014; Castañé et al., 2010; Dalton et al., 2014), latent-inhibition (Weiner et al., 1996), and attentional set-shifting (Floresco et al., 2006); our data suggest that these effects could be regulated by D1R-expressing neurons.

When interpreting the results from this set of experiments it is important to bear in mind that all rats first completed the LSQ investigating the effects of D2R antagonism with raclopride, and were later tested in a second LSQ studying the effects of D1R antagonism with SCH23390. It is possible that extra training ($\geq$ 3 reversals) could have affected animals' performance in the task by facilitating their reversal skills, or that the number of received infusions ($\geq$ 12 infusions of raclopride) could have altered the results of subsequent SCH23390 infusions. Moreover, although SCH23390 is often used to target D1R, it also has agonist affinity at the serotonin 5-HT_{2C} receptor (Millan et al., 2001), which could contribute to the observed effects in reversal learning. Nonetheless, previous studies have suggested that 5-HT_{2C} receptor manipulation in the NAc has no impact on reversal-learning performance (Boulougouris and Robbins, 2010). Perhaps more importantly, the D2R antagonist used also has strong D3R antagonism properties, which are further discussed in Chapter 7.

The present chapter adds evidence to the distinct role of DA striatal receptors in modulating reversal learning at specific stages. Together with the dorsal striatal results presented by Sala-Bayo et al. (2020), the present findings imply that visual reversal learning recruits sequential processing in ventral striatal and then dorsal striatal domains. In addition, the present results support the DA overdose hypothesis, which could explain cognitive impairments following striatal excesses of DA. These phenomenon could be differentially driven by learning from positive or negative feedback, a sensitivity that is often altered in DA-related cognitive disorders (Tremblay et al., 2002). Thus, the following chapters will investigate how different brain regions, receptors, and cellular pathways regulate learning from positive or negative feedback in reversal learning.

3.6 Conclusions

The current study elucidates the involvement of DA in reversal learning and suggests that striatal DA receptors differentially modulate this form of behavioural flexibility. Using a serial visual reversal learning task in touchscreen operant chambers, I show that infusions of D1R, not D2R, antagonist into the NAcS improves reversal learning in early perseverative stages of reversal. These results enhance our understanding of the neural circuits underlying visual reversal learning and could be relevant for cognitive inflexibility in DA-related disorders, such as Parkinson's disease (Cools et al., 2007), OCD (Denys et al., 2004) or substance use disorder (Volkow et al., 2009).

Chapter 4

Effects of dopamine D2-like receptor activation on learning from negative feedback in a reversal-learning task: systemic *versus* intra-accumbens drug administration

4.1 Introduction

DA is hypothesised to facilitate cognitive flexibility by signalling RPEs. As discussed in Chapter 1, the RPE theory postulates that midbrain DA signals the discrepancy between the expected outcome and the actual experience (Schultz, 2013; Schultz et al., 1997). As seen in Chapter 3, DA differentially modulates reversal learning depending on, not only the striatal subregion and DA receptor, but also on the reversal phase, which may inform about how animals are learning from discrepancy between reward expectancy and delivery.

In rodents and primates it has been shown that the firing rates of DA neurons within the VTA vary according to the expected reward. That is, firing rates increase in response to unexpected reward (positive RPE), and decrease in response to an unexpected lack of reward (negative RPE) or remaining unchanged if the outcome matches the prediction (Schultz

et al., 1997). This error prediction signal is thought to be processed by downstream brain regions involved in updating reward expectations and fine tuning future reward-seeking behaviour (Schultz, 2019). In support to this hypothesis, studies using *in-vivo* optogenetics show that VTA activation or inhibition bidirectionally simulate positive and negative RPEs, respectively, and affect Pavlovian reward learning (Chang et al., 2015; Steinberg et al., 2013). Experimental approaches using chemogenetics (i.e. DREADDs) have revealed that activation of the VTA to NAc pathway decreases the sensitivity of rats to losses (Verharen et al., 2018) while Klanker and colleagues (Klanker et al., 2015) observed varying levels of DA release in the NAc during different trials on a reversal learning task in rats, consistent with RPE signalling.

As discussed in Chapters 1 and 3, the model of the basal ganglia simulates how DA modulates approach and avoidance learning *via* the direct and indirect pathways of the striatum (Frank et al., 2004). However, this model has not been tested in the context of reversal learning depending on both positive and negative feedback for optimal reward outcome. It is also unknown whether impaired D2R function in the NAc is necessary and sufficient to interfere with reversal learning through effects on negative feedback (i.e. reward omission).

As described in Chapter 3, while the NAc has been demonstrated to be involved in reversal learning, findings are often equivocal. Some studies show that NAc lesions impair reversal learning performance (Dalton et al., 2014; Stern and Passingham, 1995), whereas others report no effects (Castañé et al., 2010; Schoenbaum and Setlow, 2003). This discrepancy might be due to the known anatomical and functional heterogeneity of the NAc, including the NAcC and th NAcS (Zahm, 1999), and a third region not examined in this thesis, the rostral pole. The NAcC and NAcS have been described as having distinct yet often complementary roles in reinforcement learning (Floresco et al., 2006). Supporting this functional dichotomy, inactivation of the NAcS impaired performance in a PRL task in rats, involving spurious negative feedback, whereas inactivation of the NAcC had no effect on reversal performance but instead increased latencies to respond to the stimuli (Dalton et al., 2014).

Despite studies implicating the NAc, DA afferents to this region, and D2R in reversal learning (Boulougouris et al., 2009; Radke et al., 2018), the role of D2R within subregions of the NAc remains unclear. In Chapter 3 and previous research (Sala-Bayo et al., 2020), pharmacological antagonism revealed that both NAcC and NAcS are involved in early stages of reversal learning but *via* different receptors: D1R in the NAcS and D2R in the

NAcC. However, DA depletion from the NAc reportedly impair reversal learning (Haluk and Floresco, 2009; Taghzouti et al., 1985). Microinfusions of a D2R agonist into the NAc also impaired reversal learning, and infusions of a D1R, but not D2R, antagonist in the NAc disrupted set-shifting due to increased perseveration (Haluk and Floresco, 2009). Thus, the role of D1R and D2R in the NAc – core or shell – in cognitive flexibility is ambiguous potentially due to brain regions' heterogeneity and feedback processing.

In addition, D2R are expressed both post-synaptically (i.e. in striatopallidal neurons), and pre-synaptically (i.e. in mesoaccumbal neurons). The latter act as auto-receptors that mediate auto-inhibition and suppression of DA release in conditions of high extracellular DA levels (De Mei et al., 2009). The varied synaptic location of D2R in the striatum poses challenges for the interpretation of the behavioural effects of D2R agents.

4.2 Aims, approaches, and hypotheses

In the present study, I aimed to: (1) investigate the role of D2R in approach/avoidance behaviour as learning from positive and negative feedback sensitivity during reversal learning; (2) dissociate the role of D2R in the NAcC and NAcS in regulating learning from positive or negative feedback during reversal learning; (3) distinguish whether the effects of the D2R agonist quinpirole on behaviour are mediated by pre- or post-synaptic D2R.

To address the first aim, I administered quinpirole systemically and tested visual reversal learning performance on the recently developed VPVD task in rats (see Chapter 2) (Alsiö et al., 2019). This task allowed us to discern whether animals learn from positive or negative feedback by coupling the correct or incorrect stimulus with a third, neutral stimulus that is rewarded 50% of the times. This variant of the task probes if animals learn by "approaching positive" or "avoiding negative" outcomes. Based on the Frank model above, I predicted an impairment in reversal learning caused by impaired sensitivity to negative feedback.

For the second aim, I locally infused the D2R agonist quinpirole into the NAcS and NAcC and tested visual reversal learning performance on the VPVD task, where I expected to observe the same impairment as in the systemic approach, with a potential opponent role of the two subregions.

Finally, to address the third aim, I administered systemic quinpirole to evaluate dose-dependent responses and dissociate pre- and post-synaptic D2R mediated responses. Changes in locomotor activity in a dose-response manner have been suggested to relate to pre-synaptic D2R being more sensitive to DA agonists than post-synaptic D2R (Eilam and Szechtman, 1989). I therefore hypothesised that high doses, presumably acting *via* post- (and pre-) synaptic D2R would impair reversal learning performance by impairing learning from negative feedback (Frank et al., 2004). In contrast, lower doses of quinpirole would act *via* pre- (not post-) synaptic D2R, and consequently not have the same impact on negative feedback as higher doses of quinpirole.

For experiments employing intracerebral drug infusions, I investigated the effects of an A2AR antagonist, which acts predominately on post-synaptic receptors. As described in Chapter 1, A2AR are Gs-coupled receptors and co-localise in the striatum with post-synaptic D2R in striatopallidal neurons, but not pre-synaptic D2R. Activation of A2AR stimulate the production of cAMP in neurons whereas activation of D2R inhibit the production of cAMP *via* Gi protein coupling (Fig. 1.7 in Chapter 1). Hence, I hypothesised that the effects of A2AR antagonism would resemble those of higher doses of quinpirole, which activate both pre- and post-synaptic D2R.

4.3 Material and methods

All experiments were performed at the Department of Psychology, at the University of Cambridge, UK.

4.3.1 Subjects

See Chapter 2, section 2.1, for details on housing and ethical permissions. Subjects were 133 male Lister-Hooded rats (Charles River, UK) housed in groups of four under temperature- and humidity-controlled conditions and a 12:12h dark cycle (lights on at 0700). Animals were $\sim$300 g at the beginning of behavioural procedures, and $\sim$400 g when they underwent surgery. Following surgical implantation of guide cannula, animals were single-housed.

4.3.2 Apparatus

The behavioural apparatus consisted of 28 touchscreen operant chambers (modified from Med Associates). See Chapter 2, section 2.2.1, for further details on the apparatus.

4.3.3 Drugs

For the systemic study, (-)-quinpirole hydrochloride (Sigma-Aldrich, UK), a D2R agonist, was dissolved in filtered saline to achieve doses of 0, 0.01, 0.025, 0.1, 0.25 and 0.5 mg/kg. Drugs were administered *via* the intraperitoneal (i.p.) route 60 min before testing. No adverse reactions to the repeated injections were observed.

For local infusions, the A2AR antagonist ZM-241385 (Tocris, Bioscience, UK) was dissolved in Kleptose hydroxypropyl beta-cyclodextrin (HPB) parenteral grade (Roquette UK Ltd, UK) 10% weight/volume in filtered saline and sonicated for 30 min at 30°C, which served as a vehicle. The drug was infused into the NAcC at the dose of 4 and 20 ng/μl. (-)-Quinpirole hydrochloride was dissolved in the same vehicle described above and infused into the NAcC at the doses of 0, 0.6, 6 and 20 μg/μl (*id.* 0, 0.3, 3 and 10 μg/side). Quinpirole was infused into the NAcS at 0, 0.6 and 6 μg/μl (*id.* 0, 0.3, 3 μg/side). As in the systemic approach, aliquots were frozen at -80°C in the quantities required for each testing day.

4.3.4 Behaviour

Training

Rats were initially trained following the standard training procedure for visual reversal learning. See Chapter 2, section 2.3.1, for details on behavioural training, including pre-training, visual discrimination training, and reversal learning training.

VPVD task

See Chapter 2, section 2.3.1 for details on the VPVD task and dose randomisation. In the systemic experiment, the reversal phase continued for an extra 4 days, up to 14 days in total,

to allow for asymptotic performance levels. See Fig. 2.1 in Chapter 2 for the timeline and overview of the present experimental design, as well as Table 2.1 in Chapter 2 for a summary on the VPVD task.

4.3.5 Surgeries

Bilateral 22-gauge guide cannulas (Plastics One, Sevenoaks, UK) were implanted in the NAcS (n = 29) and in the NAcC (n = 60) following standard stereotaxic techniques. See Chapter 2, section 2.4.1, for details on stereotaxic surgeries and cannula implantation. Coordinates were taken from Paxinos and Watson (1998), for the NAcS: anteroposterior (AP) +1.6, dorsoventral (DV) -1.9, mediolateral (ML) $\pm$0.75; and for the NAcC: AP +1.2, DV -1.9, ML ±1.9. All AP and ML measurements were from Bregma, whereas DV were from dura. Four metal screws and dental cement secured the guide cannulas to the skull. Obturators were introduced in the guide cannulas and protected with a dust cap.

4.3.6 Drug microinfusions

Rats implanted with guide cannulas received local microinfusions. See Chapter 2, section 2.5, for details on drug microinfusions. Injectors for microinfusions extended 5 mm below the guide cannula in both the NAcS and the NAcC, reaching a final DV of -6.9 mm from dura. Doses were pseudo-randomised depending on baseline performance on the day prior to testing and administered following a between-subject design.

4.3.7 Histologies for cannula tip placement

See Chapter 2, section 2.6, for details of the histological procedures.

4.3.8 Data analysis

The main dependent variables were percentage of correct responses (% Correct) on standard trials (A- < B+) and percentage of optimal responses (% Optimal) on probe trials (A- <

$C_{50/50}$; B+ > $C_{50/50}$). Errors from standard trials were also calculated by dividing trials into separate phases depending on levels of performance in running blocks of 30 trials (Alsiö et al., 2015). Trials were divided into "Early" or perseverative phase if animals achieved less than 11 correct trials, or "Late" if they achieved more than 19 correct trials in a block of 30 trials. Trials in between both criteria were classified as "Mid" or the random phase. Latencies to collect the reward and to respond to the stimuli were averaged across sessions. For the systemic drug data, I additionally examined the total number of trials. Latencies were also analysed by splitting them into each type of trial: correct (i.e. choosing B+ over A-), incorrect (i.e. choosing A- over B+), optimal (i.e. choosing B+ over $C_{50/50}$), suboptimal (i.e. choosing A- over $C_{50/50}$) or probe (i.e. choosing $C_{50/50}$ over A- or B+). In relation to reward collection latencies, only rewarded trials were analysed.

See Chapter 2, section 2.7, for further details on statistical hypothesis testing. To ensure normality, responses were arcsine-transformed; errors for each per phase square-root transformed, and latencies $\log_{10}$-transformed. Choice data were then subjected to Linear Mixed-Effects Model analysis using the lmer package in R. The model contained two fixed factors (dose, session) and one factor (subject) modelled as a random slope to account for individual differences between rats across phase (i.e. individual learning curves). Analysis of latencies involved two fixed factors (dose, latency) and subjects as random variability. When latencies were collapsed over sessions, data were subjected to a one-way ANOVA with one within-subject factor (dose). Errors per phase analysis included two fix factors (dose, phase) and the subject factor to account for individual differences. When significant dose $\times$ session interactions were found, analysis was followed by *post hoc* Dunnett's corrected pairwise comparisons.

4.4 Results

4.4.1 Histology

The ventral-most locations of injectors are shown in Fig. 4.2B. Rats were excluded from the study if the injectors cannulas were positioned outside the target areas (n = 3 NAcC; n = 2 NAcS). Targeted coordinates and final group sizes are shown in Fig. 4.1A. Each rat underwent a single dose treatment, as experiment was based on a between-subject design.

A)

	NAcS	NAcC
Guide cannulas		
AP	+1.60	+1.20
ML	±0.75	±1.90
DV	-1.90	-1.90
Injectors		
DV	-6.90	-6.90
n		
Vehicle	8	15
Q0.3 µg/side	8	12
Q3 µg/side	5	11
Q10 µg/side	N/A	7
Z4 ng/µl	N/A	6
Z20 ng/µl	8	9
Total	**29**	**60**

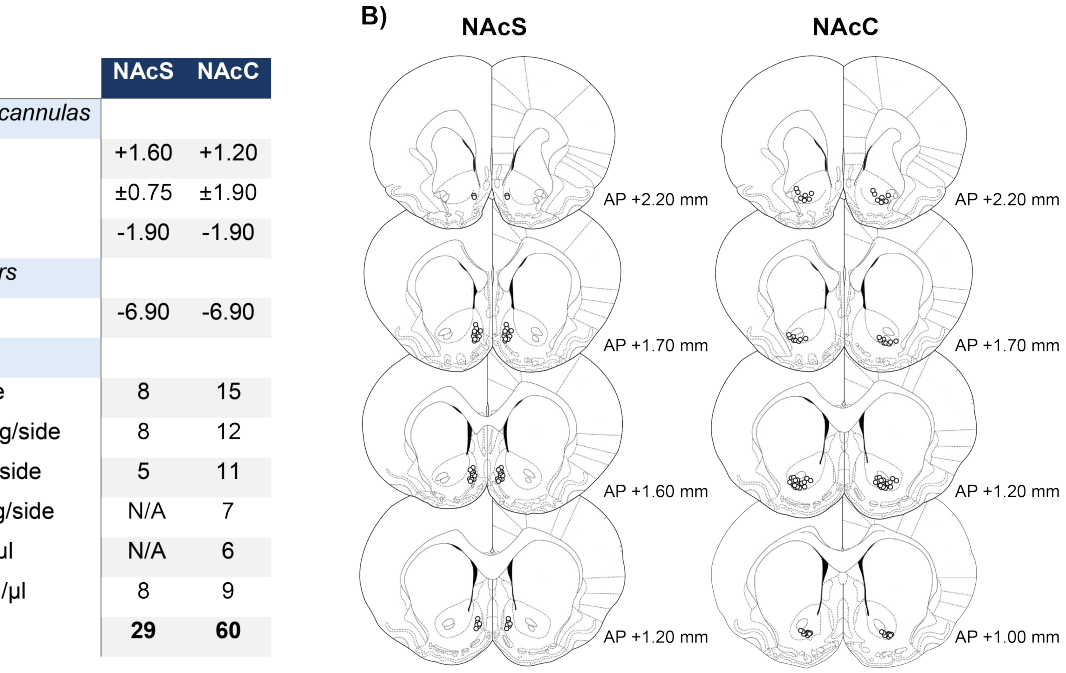


Fig. 4.1 A) Coordinates and group sizes for the injector placements in shell and core of the nucleus accumbens (NAcS; NAcC). Anteroposterior (AP) and mediolateral (ML) coordinates were measured from Bregma and dorsoventral (DV) from dura. B) Ventral-most injector tip placements in the NAcS and the NAcC. C) Example of bilateral injector tip placement in the NAcS and NAcC.

4.4.2 Behavioural results

Systemically administered quinpirole dose-dependently impairs reversal learning by blocking learning from negative feedback

Systemic administration of quinpirole impaired overall reversal learning performance by hindering learning from negative-probe trials, which led to reduced avoidance of the non-rewarded stimulus (Fig. 4.2).

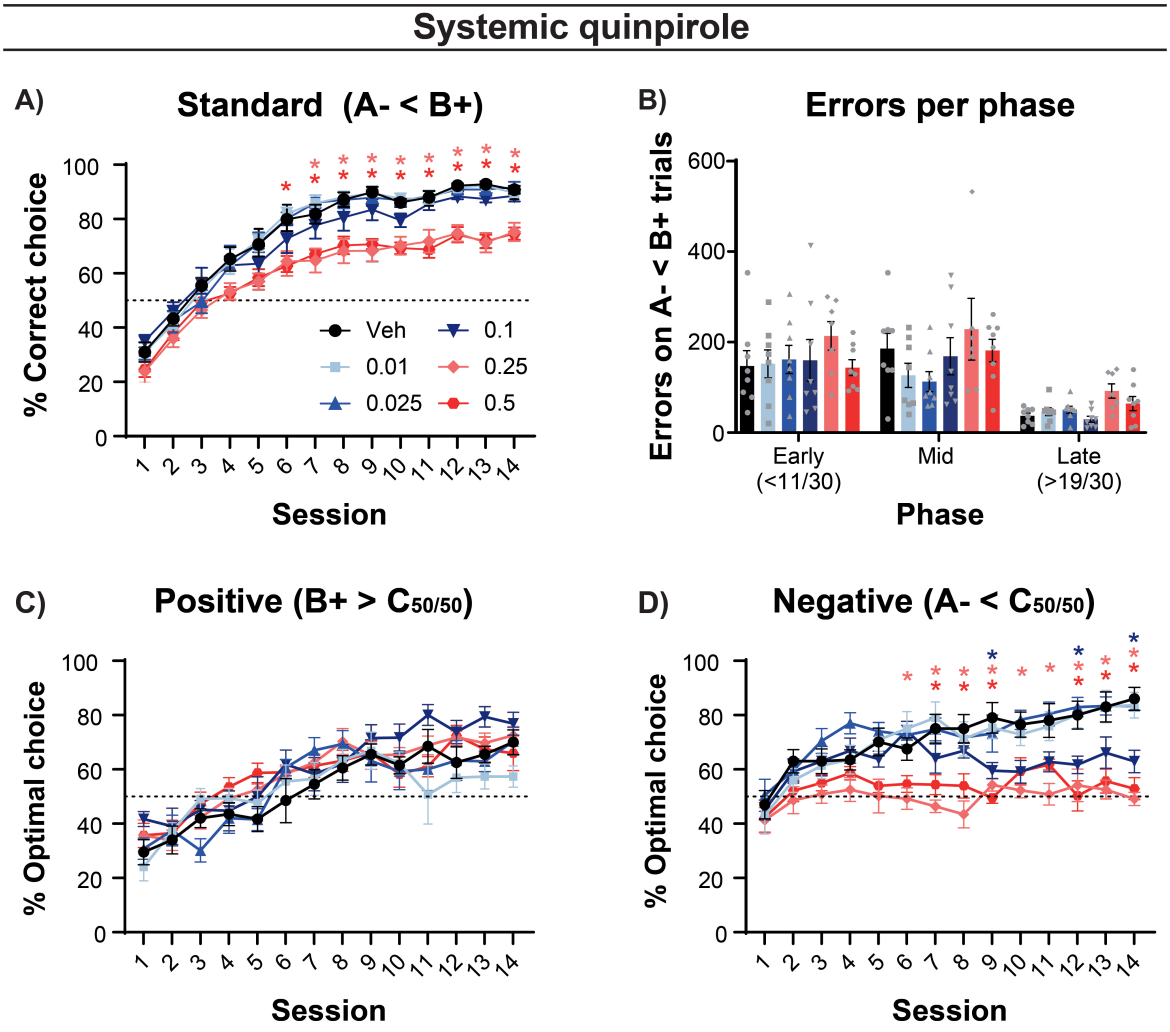


Fig. 4.2 The DA D2R agonist quinpirole impaired visual reversal learning in the VPVD task by impairing learning from negative feedback. A) Standard reversal learning trials (A- < B+). Quinpirole at high doses (0.25 and 0.5 mg/kg) impaired performance by decreasing the percentage of correct responses. B) Errors on standard trials per phase: early (< 11 correct responses in blocks of 30 trials), mid, late (> 19 correct responses in blocks of 30 trials). Data were collapsed throughout sessions. Quinpirole did not affect the number of errors per phases. C) Positive-probe trials (B+ > $C_{50/50}$). Quinpirole did not alter the percentage of optimal responses. D) Negative-probe trials (A- < $C_{50/50}$). Quinpirole dose-dependently decreased percentage of optimal choices, indicating an impairment in learning from negative feedback. Data are shown as mean ± SEM. * $p < 0.05$ *vs* vehicle-treated group.

On standard trials (Fig. 4.2A), linear mixed-effects model showed a main effect of dose ($F_{5, 41} = 9.903$, $p < 0.001$) and a significant session × dose interaction ($F_{65, 533} = 1.571$, $p = 0.004$). *Post-hoc* analysis revealed a significant difference with dose 0.25 mg/kg from session 7 onwards in comparison to the vehicle-treated group (all days p = 0.002 - 0.041); and dose 0.5 mg/kg from session 6 onwards (all days p = < 0.001 – 0.038), indicating impaired performance from mid to late stages.

When errors were analysed splitting trials into early (< 11 correct in 30 trials), mid, and late (> 19 correct in 30 trials) phases (Fig. 4.2B), a significant effect of phase was revealed ($F_{2, 82} = 30.23$, $p < 0.001$), but there was no effect of dose ($F_{5, 41} = 1.948$, $p = 0.107$) or phase × dose interaction ($F_{10, 82} = 0.564$, $p = 0.836$). Hence, changes in phases did not depend on treatment.

Within negative trials (Fig. 4.2D), there was a main effect of dose ($F_{5, 41} = 11.90$, $p < 0.001$) and a significant session × dose interaction ($F_{65, 533} = 2.212$, $p < 0.001$). *Post-hoc* comparisons revealed a significant effect with the high doses: 0.1 mg/kg during session 9 and 12 ($p = 0.042$, $p = 0.041$), 0.25 mg/kg from session 6 onwards (all p = < 0.001 - 0.028), and dose 0.5 mg/kg from session 7 to 9 and from session 12 to 14 (all p = < 0.001 - 0.032). On positive trials (Fig. 4.2C), no effect of dose ($F_{5, 41} = 1.792$, $p = 0.136$) or session × dose interaction ($F_{65, 533} = 1.293$, $p = 0.070$) were observed. Probe trials therefore showed that animals were impaired in reversal learning selectively due to processing negative feedback from mid to late stages.

The total number of trials achieved per session were also averaged and analysed (Table 4.1A). One-way ANOVA revealed an effect of dose ($F_{5, 41} = 4.592$; $p = 0.002$). *Post-hoc* analysis showed that only the highest dose decreased the number of trials completed per session in comparison with the vehicle control group (0.5 mg/kg; $p = 0.001$).

Quinpirole treatment dose-dependently increased the latency to collect the reward ($F_{5, 41} = 13.79$, $p < 0.001$; Table 4.1A), but had no effect on the latency to respond to the screen ($F_{5, 41} = 0.188$, $p = 0.766$; Table 4.1A). *Post-hoc* analysis showed that the doses of 0.025, 0.1, 0.25 and 0.5 mg/kg increased overall latencies to respond to the stimuli (p = 0.044, < 0.001, < 0.001, < 0.001, respectively).

Systemic quinpirole

A)

Variable	Doses (mg/kg)	Latencies to respond (ms)	Latencies to collect (ms)	Trials per session
Latencies collapsed across trials	Vehicle	1066 ± 38	1529 ± 50	197.99 ± 1.77
	0.01	1231 ± 76	1786 ± 114	194.34 ± 3.81
	0.025	1261 ± 119	*1916 ± 133**	193.96 ± 3.34
	0.1	1179 ± 59	*2257 ± 104****	191.93 ± 3.60
	0.25	1183 ± 126	*2559 ± 118****	190.97 ± 3.61
	0.5	1213 ± 93	*2556 ± 102****	*168.15 ± 9.81****

B)

Latencies (ms)	Doses (mg/kg)	Standard B+ over A-	Standard A- over B+	Probe B+ over $C_{50/50}$	Probe A- over $C_{50/50}$ **	Probe $C_{50/50}$ over A- or B+
To respond per type of trials	Vehicle	1135 ± 47	1156 ± 53	1110 ± 70	*1174 ± 67*	1122 ± 55
	0.01	1174 ± 85	1161 ± 85	1187 ± 97	*1368 ± 154*	1230 ± 138
	0.025	1171 ± 83	1214 ± 139	1167 ± 97	*1275 ± 104*	1141 ± 106
	0.1	1104 ± 76	1021 ± 42	1056 ± 87	*1166 ± 58*	1078 ± 87
	0.25	1302 ± 137	1387 ± 156	1253 ± 132	*1434 ± 161*	1300 ± 130
	0.5	1280 ± 108	1265 ± 131	1189 ± 108	*1330 ± 126*	1293 ± 98

C)

Latencies (ms)	Doses (mg/kg)	Standard B+ over A-	Probe B+ over $C_{50/50}$	Probe $C_{50/50}$ over A-
To collect per type of trials	Vehicle	2230 ± 192	2264 ± 165	2219 ± 208
	0.01	2268 ± 223	2291 ± 220	2305 ± 220
	0.025	1952 ± 165	1962 ± 171	1943 ± 165
	0.1	2055 ± 165	1916 ± 175	2140 ± 186
	0.25	1929 ± 163	2062 ± 157	1943 ± 161
	0.5	2078 ± 155	2078 ± 168	2058 ± 165

Table 4.1 Following systemic administration of quinpirole, latencies to respond to the visual stimuli and to collect the sucrose pellet from the magazine. A) Latencies and trials per session collapsed throughout sessions. Latencies to collect the reward following administration of quinpirole at doses 0.025, 0.1, 0.25 and 0.5 mg/kg were higher than those from the vehicle-treated group were. Trials per session decreased with quinpirole at dose 0.5 mg/kg. B) Latencies to respond to the stimuli were split depending on the type of trial. Latencies on negative-probe trials in which animals made an error were longer than the other latencies. C) Latencies to collect the reward were segregated according to type of trials. Quinpirole did not affect latencies to collect the reward. Data are presented as mean ± SEM, collapsed across sessions. * $p < 0.05$, *** $p < 0.001$ *vs* vehicle-treated group; ** $p < 0.01$ *vs* other trials latencies.

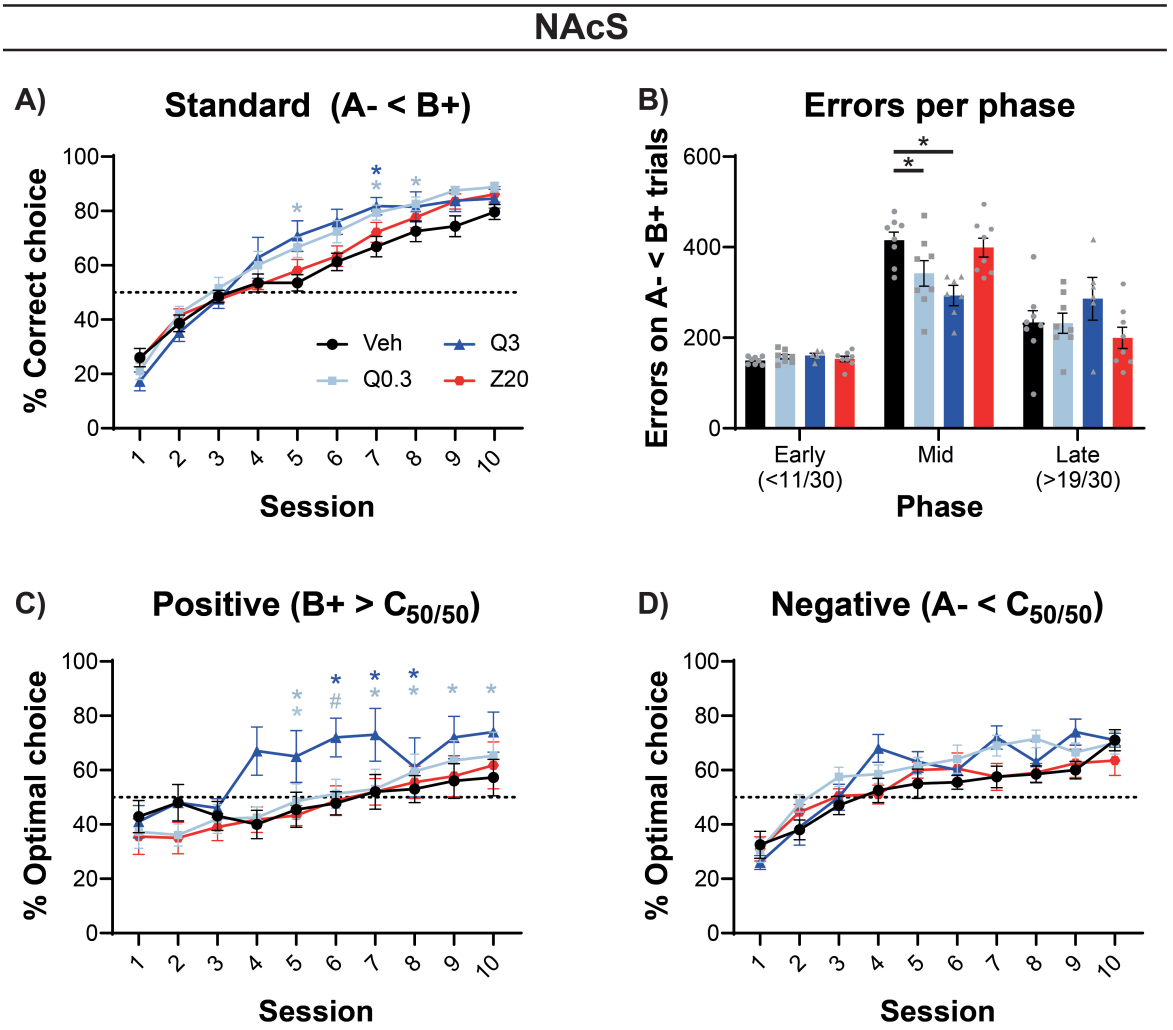


Fig. 4.3 Improved visual reversal learning by intra-NAcS infusions of the D2R agonist quinpirole manifests from accentuated learning from positive feedback. A) Standard reversal learning trials (A- < B+). Quinpirole at 0.3 and 3 μg/side improved performance. B) Errors on standard trials split by reversal-learning phase: early (< 11 correct responses in sets of 30 trials), mid and late (> 19 correct responses in sets of 30 trials). Quinpirole at 0.3 and 3 μg/side decreased the number of errors during the mid-phase of reversal learning. Trials were collapsed throughout sessions. C) Positive-probe trials (B+ > $C_{50/50}$). Quinpirole at 0.3 and 3 μg/side increased optimal responses, indicating an improvement in learning from positive feedback. D) Negative-probe trials (A- < $C_{50/50}$). Neither quinpirole nor ZM-241385 affected percentage of optimal choices, indicating intact learning from negative feedback. Data are shown as mean ± SEM. $^{\#}$ p ~0.05, * p < 0.05 *vs* vehicle-treated group.

Latencies were next analysed depending on the trial type i.e. on standard trials choosing B+ over A-, or *vice versa*; and on probe trials choosing B+ over $C_{50/50}$, A- over $C_{50/50}$, or $C_{50/50}$ over B+ or A- (Table 4.1B). On latencies to respond, I found a main effect of type of trial ($F_{4, 65} = 8.026$, $p < 0.001$), but no significant dose $\times$ type of trial interaction ($F_{20, 164} = 0.695$, $p = 0.827$). *Post-hoc* analysis revealed that the difference was driven by those trials in which animals chose the incorrect stimulus (A-) over the neutral stimulus ($C_{50/50}$), so when animals made an error in negative-probe trials (A- over $C_{50/50}$ *vs* B+ over A-, $p = 0.001$; *vs* A- over B+, $p = 0.001$; *vs* B+ over $C_{50/50}$, $p < 0.001$; *vs* $C_{50/50}$ over either A- or B+, $p < 0.001$). On latencies to collect the reward depending on each type of trial, I found no effect of dose ($F_{5, 41} = 0.517$, $p = 0.762$) or dose $\times$ type of trial interaction ($F_{10, 82} = 1.533$, $p = 0.142$) (Table 4.1C).

Intra-NAcS infusions of quinpirole but not ZM-241385 improves reversal learning

Following infusions into the NAcS (Fig. 4.3), the D2R agonist quinpirole improved reversal-learning performance by increasing learning from positive feedback. The A2AR antagonist ZM-241385 had no effect.

On standard trials (A- < B+), I found a significant dose $\times$ session interaction ($F_{3, 237.71} = 9.002$, $p < 0.001$) (Fig. 4.3A). *Post-hoc* analysis revealed that the 0.3 μg/side dose of quinpirole increased the percentage of correct responses from session 5 until session 9. The dose of 3 μg/side had the same effect from sessions 5 to 7. In contrast, ZM-241385 did not differ from the vehicle control group in any session of the VPVD task, thus it did not replicate quinpirole results.

Quinpirole selectively improved performance from positive feedback while leaving intact learning from negative feedback. On the positive-valence probe trials (B+ > $C_{50/50}$), the Linear Mixed-Effects Model revealed an effect of dose ($F_{3, 247.31} = 5.319$, $p = 0.001$) and dose $\times$ session interaction ($F_{3, 232.04} = 5.398$, $p = 0.001$). *Post-hoc* comparisons showed that this effect was specific for quinpirole 0.3 μg/side from session 5 onwards, and 3 μg/side between sessions 5 and 6, respectively (Fig. 4.3C). No changes were detected on the negative-valence probe trials (A- < $C_{50/50}$) (dose: $F_{3, 248.08} = 1.471$, $p = 0.223$; dose $\times$ session: $F_{3, 247.53} = 1.535$, $p = 0.206$) (Fig. 4.3D).

On errors during standard trials when trials were split into early, mid and late phases (Fig. 4.3B), I found a significant dose $\times$ phase interaction ($F_{6, 75} = 2.674$, $p = 0.021$). Further analysis showed a significant main effect of dose in the mid phase ($F_{3, 25} = 4.273$, $p = 0.015$).

Post-hoc analysis revealed this effect was driven by the two doses of quinpirole in comparison to the vehicle group: 0.3 μg/side ($p = 0.025$) and 3 μg/side ($p = 0.025$).

High doses of quinpirole, though not ZM-241385, selectively increased latencies to collect the reward, while did not alter time to respond to the stimuli. Analysis of latency to collect the reward showed an overall increase regardless of the type of trials (main effect of dose: $F_{3, 75} = 12.304$, $p < 0.001$; dose $\times$ type of trial interaction: $F_{6, 75} = 0.075$, $p = 0.998$). *Post-hoc* contrasts determined that this effect was apparent for the 0.3 and 3 μg/side doses of quinpirole (both $p < 0.001$). Similarly, in relation to latencies to respond, I observed a main effect of dose ($F_{3, 125} = 10.855$, $p = < 0.001$), but not a significant dose $\times$ type of trials interaction (Ftextsubscript12, 125 = 0.250, p = 0.995). The main effect of dose was driven by the effect of quinpirole at doses 0.3 and 3 μg/side ($p = 0.007$; $p < 0.001$) (Table 4.2).

Experiment	Doses	Latency to respond (ms)	Latency to collect (ms)
NAcS	Vehicle	1246.44 ± 168.89	1421.15 ± 59.43
	Q 0.3 µg/side	1537.96 ± 173.82	2082.64 ± 222.54
	Q 3 µg/side	1972.35 ± 361.36$^{\#}$	*2266.85 ± 387.54**
	Z 20 µg/µl	1365.88 ± 105.23	1610.89 ± 134.51
NAcC	Vehicle	1919.09 ± 281.00	1773.96 ± 97.47
	Q 0.3 µg/side	1564.43 ± 146.14	1862.73 ± 116.18
	Q 3 µg/side	1632.06 ± 117.61	*2157.94 ± 130.77**
	Q 10 µg/side	1938.41 ± 132.67	*2276.77 ± 105.28**
	Z 4 µg/µl	1467.20 ± 253.37	1649.07 ± 214.49
	Z 20 µg/µl	1218.63 ± 90.61	1610.24 ± 101.54

Table 4.2 Following microinfusions of drugs, latencies to respond at the screen and to collect the sucrose pellet from the magazine on rewarded trials. Data are shown as mean ± SEM, collapsed across sessions. $^{\#}$ p ∼0.05, * $p < 0.05$ *vs* vehicle-treated group. Quinpirole (Q), ZM-241385 (Z).

Intra-NAcC infusions of quinpirole or an ZM-241385 affect reversal learning

Quinpirole at low doses and ZM-241385 both improved reversal-learning performance in the mid-stages by improving learning from negative feedback. In contrast, the highest dose of quinpirole (10μg/side) impaired performance in reversal learning by decreasing learning from negative feedback during the late stages (Fig. 4.4).

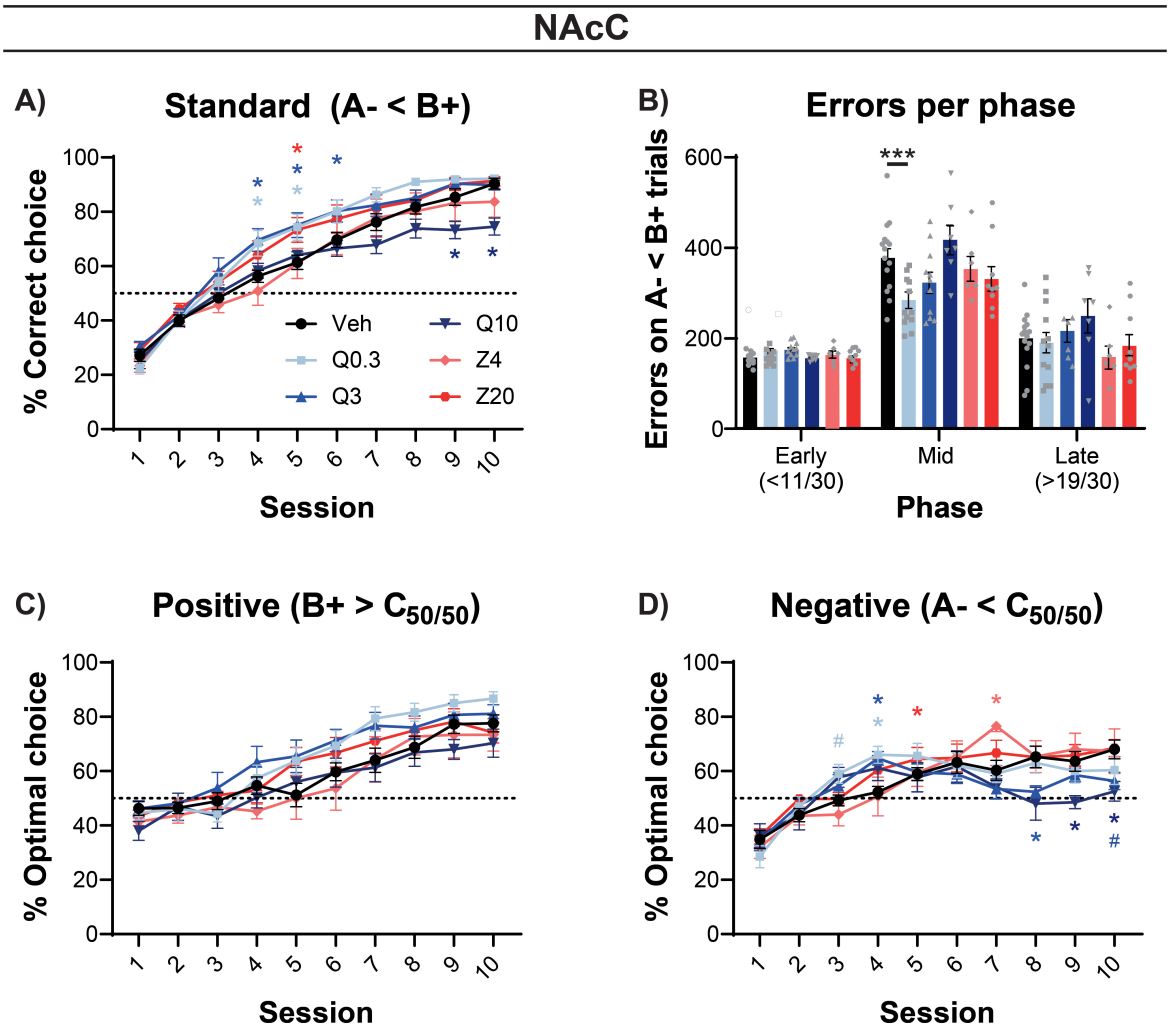


Fig. 4.4 Low doses of the D2R agonist quinpirole in the NAcC improves visual reversal learning in the VPVD task whereas a higher dose of quinpirole impairs reversal learning. A) Standard reversal learning trials (A- < B+). Quinpirole at 0.3 and 3 μg/side and ZM-241385 at 20 ng/µl improved, whereas quinpirole at 10 μg/side impaired performance. B) Errors on standard trials split by reversal-learning phase: early (< 11 correct responses in sets of 30 trials), mid and late (> 19 correct responses in sets of 30 trials). Quinpirole at 0.3 μg/side decreased the number of errors on standard trials during the mid phase of reversal learning. Trials were collapsed throughout sessions. C) Positive-probe trials (B+ > $C_{50/50}$). Drugs did not affect optimal responses. D) Negative-probe trials (A- < $C_{50/50}$). Quinpirole at 0.3 and 3 μg/side increased optimal responses, indicating an improvement in learning from negative feedback. Quinpirole at 10 μg/side decreased optimal responses at later sessions, indicating an impairment in learning from negative feedback. Data are shown as mean ± SEM. $^{\#}$ p $\sim$0.05, * p < 0.05, *** p < 0.001 *vs* vehicle-treated group.

On standard trials (A- < B+) (Fig. 4.4A), I found a main effect of dose ($F_{5, 54} = 3.188$, $p = 0.014$) and a dose × session interaction ($F_{45, 485} = 2.043$, $p < 0.001$). *Post-hoc* analysis with Dunnett's correction showed that the dose of quinpirole 0.3 μg/side increased percentage of correct responses in comparison to vehicle in sessions 4 ($p = 0.016$) and 5 (0.009). This same effect was observed with the dose of 3 μg/side of quinpirole ($p = 008$; $p = 0.006$), and with ZM-241385 in session 5 ($p = 0.038$). The dose of 10 μg/side decreased the percentage of correct responses in session 10 ($p = 0.006$), with a trend decease during session 9 ($p = 0.060$). Thus, quinpirole and ZM-241385 improved reversal learning, potentially during mid-stage sessions.

When the number of errors during standard trials was analysed by splitting the trials into early (< 11 correct/30 trials), mid or late (> 19 correct/30 trials) phases (Fig. 4.4B), a significant dose × phase interaction was found ($F_{10, 108} = 2.563$, $p = 0.022$). Further analysis revealed a change in errors in the mid phase of reversals learning (main effect of dose: $F_{5, 54} = 4.858$, $p = 0.001$), which *post-hoc* analysis suggested was driven by the quinpirole dose of 0.3 μg/side ($p = 0.001$), confirming the quinpirole-induced improvement was specific to the mid phase.

Analysis of the choice performance on probe trials showed that quinpirole- and ZM-241385-affected reversal learning was the result of selectively altering behaviour on negative-probe trials (Fig. 4.4D). I observed a significant dose × session interaction ($F_{45, 485} = 2.108$, $p < 0.001$). *Post-hoc* analysis showed that the effect on negative trials was driven by an increase of optimal choices in session 4 following administration of quinpirole at 0.3 μg/side ($p = 0.007$) and 3 μg/side ($p = 0.005$), and in session 7 following administration of ZM-241385 at dose 4 ng/μl ($p = 0.005$). There was also a decrease in optimal choices after quinpirole at dose 3 μg/side during session 8 ($p = 0.04$) and at 10 μg/side from session 9 onwards ($p = 0.011$; $p = 0.031$). In the last session, the 3 μg/side dose of quinpirole appeared to also decrease the percentage of optimal responses but this did not reach statistical significance ($p = 0.083$). No effect was observed on positive-probe trials (dose × session: $F_{45, 485} = 1.129$, $p = 0.267$), although there was a trend for a main effect of dose ($F_{5, 54} = 2.142$, $p = 0.074$) (Fig. 4.4C). All tested drugs and doses left optimal choice in positive-probe trials intact, but high doses of quinpirole impaired choice on negative-probe trials during mid to late stages.

Latencies to collect the reward or to respond to the stimuli were analysed with reference to the type of trial they were measured from: A- < B+, B+ > $C_{50/50}$ and A- < $C_{50/50}$, and for responding latencies when they made an incorrect response too, i.e. choosing A- over

$C_{50/50}$ or $C_{50/50}$ over B+. Although I did not observe a significant dose $\times$ type of latency for latencies to collect the reward ($F_{10,\ 534} = 0.048$, $p = 0.998$) or to respond ($F_{20,\ 534} = 0.308$, $p = 0.988$), there was a main effect of dose with respect to latencies to collect the reward ($F_{10,\ 534} = 3.869$, $p = 0.016$). *Post-hoc* analysis showed that quinpirole at doses 3 ($p = 0.043$) and 10 μg/side ($p = 0.022$) increased the time to collect the reward from the magazine (Table 4.2). Thus, quinpirole doses that affected performance by altering performance on responding to negative feedback also increased selectively the time to collect the reward, while did not affect animals' speed to respond to the stimuli.

Overall, these experiments showed dissociable roles of D2R modulating reversal learning performance when being quinpirole-activated across all brain regions by systemic injections or locally into the NAcC and NAcS. Systemic quinpirole impaired reversal-learning performance by decreasing avoidance of the negative stimulus, whilst local infusions into the NAc generally accelerated learning by promoting approach to the positive stimulus (modulated by the NAcS), or increasing avoidance of the negative stimulus (modulated by the NAcC). In addition, an impairment in reversals was observed with the highest dose of quinpirole into the NAcC.

4.5 Discussion

The present study demonstrated the theoretical advantages of using the recently developed VPVD task, specifically in dissociating how learning from positive and negative feedback influences visual reversal learning. I found strong supporting evidence for the hypothesis that D2R activation impairs reversal learning by blocking learning from losses when targeted systemically. In contrast, I found dissociable effects when quinpirole was administered locally into the NAcS and NAcC, mainly leading to an improvement in reversal learning by increasing approach responses to rewarded stimuli (NAcS) or increasing avoidance responses to negative feedback (NAcC). With respect to the NAcC, I also found blunted learning from negative feedback following administration of high doses of quinpirole.

Systemic D2R agonism with quinpirole impairs visual reversal learning by blunting learning from negative feedback

Systemic administration of quinpirole severely disrupted reversal learning at the highest administered doses, while left learning from positive feedback intact. This impairing effect

is in agreement with my hypothesis and previous studies where D2R agonism impaired reversal learning in rats (Boulougouris et al., 2009), non-human primates (Smith et al., 1999), and humans (Mehta et al., 2001). The present results extend such findings by revealing a dose-dependent and highly selective effect of quinpirole on negative-probe trials in the VPVD task, indicating a selective deficit on learning from negative feedback. Rats treated with high doses of quinpirole showed poorer performance on standard trials than rats treated with lower doses or vehicle, and remained at random levels on negative trials, even after two weeks of training.

To further explore these findings, Alsiö and colleagues (Alsiö et al., 2019) investigated the effects of quinpirole on performance of the PRL task. The results of this experiment concurred with the findings from the novel VPVD task by revealing a dose-dependent decrease in the number of reversals per session. Importantly, the PRL study was complemented by a computational model of the behavioural results based on Rescorla-Wagner reinforcement learning (Rescorla and Wagner, 1972) and hierarchical Bayesian analysis (Daw, 2009). The model showed that the highest dose of quinpirole (0.25 mg/kg) selectively decreased the learning rate from losses (i.e. α_{loss} parameter), representing learning from negative feedback, whilst left learning from positive feedback intact (i.e. α_{win} parameter) ((Alsiö et al., 2019); Fig. 4.5). The computational analysis also revealed enhanced inverse temperature (i.e. β parameter), the proportion of how much subjects "exploit" or "explore" the stimuli depending on expected reward properties (Alsiö et al., 2019). Higher β parameters indicate an increase in reinforcement sensitivity and that animals are driven by exploiting the choice they expect is going to be rewarded. In a reversal situation, anticipated reinforcement is placed on the stimulus that was previously rewarded (now unrewarded), so in the absence of exploration of the previous unrewarded (now rewarded) stimulus, animals would perform suboptimally. According to this logic, therefore, high dose quinpirole may have impaired reversal learning by attenuating learning from losses and/or by increasing exploitation of the non-rewarded stimulus.

A wide range of doses was used to determine if quinpirole-induced effects were mediated by pre- or post-synaptic D2R. As hypothesised, high doses of quinpirole impaired reversal performance, suggesting the drug is modulating behaviour acting *via* post- (and pre-) synaptic D2R (Frank et al., 2004). It is worth noting that whereas the higher doses affected performance on reversal (> 0.1 mg/kg), the dose of 0.5 mg/kg not only modified percentage of correct and optimal choices, but also the number of trials completed per session. The observed impairment in flexible behaviour was similar to the effects of 0.25 mg/kg quinpirole,

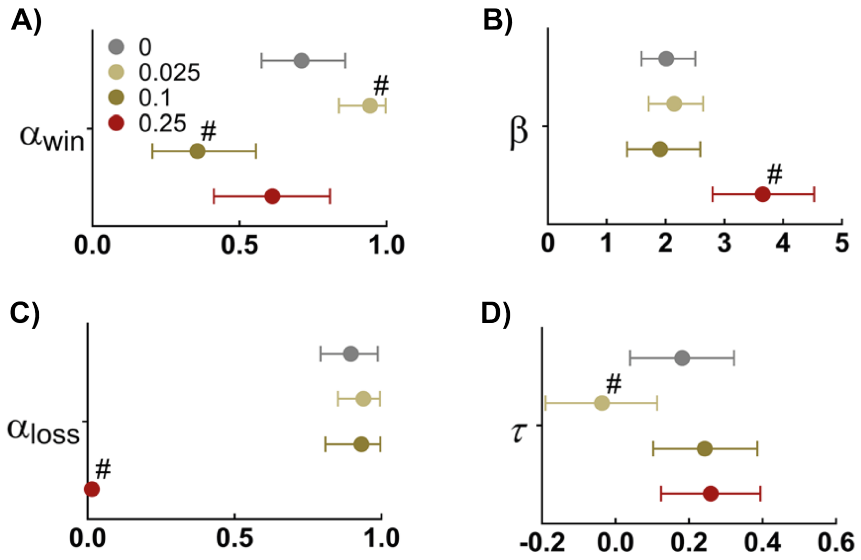


Fig. 4.5 Computational model parameters estimated by hierarchical Bayesian analysis of trial-by-trial choice data from the PRL task. D2R agonist quinpirole impaired learning rate after losses and increased the increase temperature or likelihood to explore over exploit a stimulus. A) Learning rate from rewards (α_{win}). B) Inverse temperature or exploit/explore ratio (β). C) Learning rate from lack of reward (α_{loss}). D) Stickiness parameter or likelihood to repeat the choice from the previous trial regardless of its rewarding properties (τ). From Alsiö et al. (2019).

but we cannot exclude the possibility that rats were slower to learn with 0.5 mg/kg quinpirole due to a reduced opportunity to sample the reward contingencies of the stimuli (Table 4.1A).

Dissociable effects of subregion specific infusions of quinpirole on reversal learning

Contrary to the effects of systemically administered quinpirole on reversal learning, the administration of quinpirole to restricted subregions of the NAc generally improved reversal-performance. When infused into the NAcS, quinpirole at doses 0.3 and 3 μg/side improved performance, mainly by increasing optimal choices on positive-probe trials, indicating an enhanced sensitivity to positive feedback. The A2AR antagonist, ZM-241385, infused into this same region had no significant effect on reversal learning compared with the vehicle-treated group. Conversely, when infused into the NAcC, quinpirole at doses 0.3 and 3 μg/side improved performance during the mid-stages of reversal by improving performance on negative-probe trials, which was also observed following the administration of the A2AR antagonist. In addition, the highest dose of quinpirole into the NAcC impaired performance during late sessions by blunting learning from negative feedback.

In relation to the NAcC, I obtained a complex set of findings, which was partly at odds with my hypotheses. The lower doses of quinpirole improved reversal-learning performance on standard trials, as did the A2AR antagonist, ZM-241385, by selectively improving optimal choices on negative-probe trials. However, at the highest dose of 10 μg/side, quinpirole impaired reversal-learning performance and decreased optimal responses on negative-probe trials, indicative of blunted learning from negative feedback. The impairing effects of 10 μg/side quinpirole is broadly consistent with my hypothesis and similar effects of systemically administered quinpirole. This deficit perhaps accords with quinpirole acting at post-synaptic D2R but surprisingly was not matched by ZM-241385 (Fig. 4.6 for visual interpretation), which instead resembled the enhancing effect of lower dose quinpirole. Unfortunately, it is not possible to definitively demonstrate whether the behavioural effects were the result of activating pre-synaptic D2R (or D2R on interneurons), or the implication of a different mechanism, including the involvement of other brain regions such as the dorsal striatum. Nonetheless, while lower doses of quinpirole lost their enhancing effect on negative-probe trials and performance returned to vehicle level in sessions 5 and 6, in later sessions, the dose of 3 µg/side decreased performance even below the control group, suggesting that such dose is intermediate both in terms of dose range and effects. This suggests the presence of a potential adaptive mechanism to high or repeated doses of quinpirole, possibly as the result of sensitisation to the drug or internalisation of D2R. Thus, the effect of the highest dose of quinpirole could be regulating behaviour *via* diminishing the role of post-synaptic D2R. It would be of interest to repeat these experiments including the dose of 10 μg/side into the NAcS to determine if the impairment is exclusively modulated by the NAcC in the ventral striatum or if the NAcS also plays a role in the deficit.

In contrast, lower doses of quinpirole improved performance by activating post-synaptic D2R, as suggested by the replication of these effects with ZM-241385 (Fig. 4.6C). This result did not match my hypothesis of an impairment driven by post-synaptic receptors. The discrepancy between my results (i.e. improvement when overactivating D2R) and what Frank's model of the basal ganglia would have predicted (i.e. impairment when overactivating D2R) could be relate to a less dichotomous dissociation of the direct and indirect pathways expressing D1R and D2R in the ventral striatum in comparison to the dorsal striatum (Kupchik et al., 2015). Hence Frank's model potentially being more accurate for the dorsal striatum. Indeed, up to one-third of neurons in NAc have been suggested to express both D1R and D2R together on the same MSN (Cole et al., 2018; Humphries and Prescott, 2010); Fig. 1.4 Chapter 1). Also, both pathways extend inhibitory projections towards the

other pathway that are sufficient to produce lateral inhibition (Salery et al., 2020), which may alter neuronal activity. Hence, the observed modulation could result from combined D1R- and D2R-expressing MSNs activity in the NAcC (Fig. 4.6D). In addition, the A2AR antagonist ZM-241385 was used as a probe to dissociate pre- *versus* post-synaptic effects of quinpirole (Fig. 4.6B). While I expected ZM-241385 to replicate quinpirole results if this was modulating behaviour *via* post-synaptic D2R, hence when high doses of quinpirole were infused. However, ZM-241385 matched the effect of lower doses of quinpirole. While other receptors or brain regions could be involved, as discussed, it is important to bear in mind when interpreting the results that doses of quinpirole and antagonising A2AR constitute indirect approaches to assess receptor specificity, and we cannot exclude the involvement of other mechanisms, especially when the ZM-241285 did not differ from the vehicle control group. Clear definitive evidence could be sought by producing pathway-specific D2R knockdowns and comparing the results with drug infusions, or selectively manipulating these pathways with an inhibitory opsin or the expression of Gi-coupled DREADDs (Deisseroth, 2011; Nichols and Roth, 2009). Chapter 6 describes an attempt to use optogenetic approaches to investigate another reversal paradigm.

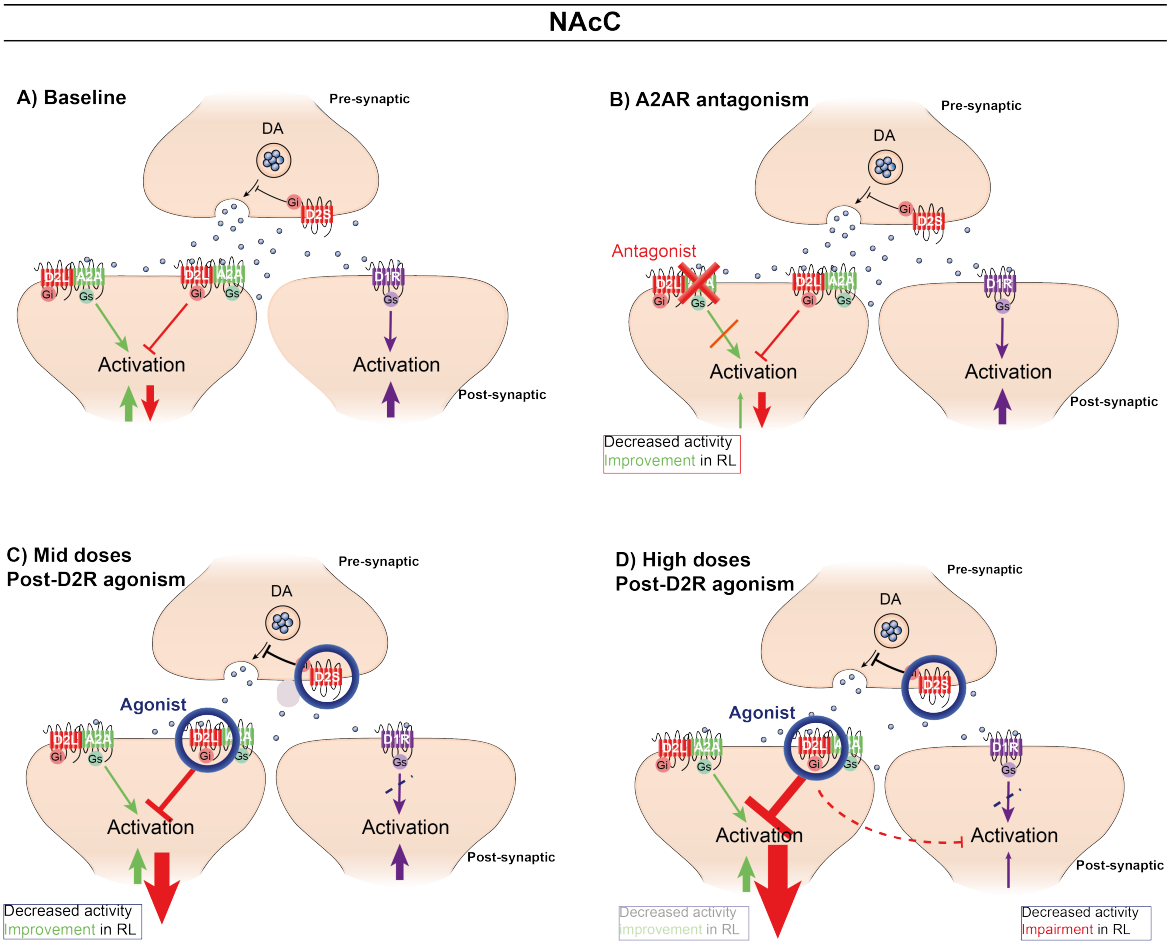


Fig. 4.6 Representation of the potential effects of D2R agonism and A2AR antagonism in NAcC MSNs. A) Baseline signalling by D1R- and D2R-expressing MSNs. DA is released by pre-synaptic neurons that express D2S, the short isoform of D2R, as auto-receptors. DA can then join D1R and D2R (mainly the long isoform, D2L) in post-synaptic neurons and activate their signalling cascade. A2AR are co-expressed with D2R in post-synaptic neurons. B) Interpretation of A2AR antagonism by ZM-241385. The antagonist may join A2AR and block their downstream signalling, resulting in lower cell activation, and decreased pathway signalling, which improved reversal learning performance by enhancing learning from negative feedback. C) Interpretation of D2R agonism by mid doses of quinpirole. As the effects matched those induced by the post-synaptic D2R-probe agent, ZM-241385, mid doses of quinpirole into the NAcC might be acting via post- (as well as pre-) synaptic D2R. Agonising these receptors leads to increased inhibition of cell activation, hence reduced signalling via efferent D2R-expressing MSNs. This improved reversal learning performance by enhancing learning from negative feedback. D) Interpretation of D2R agonism by high doses of quinpirole. Increased post-synaptic D2R agonism might lead to further regulatory mechanisms, including lateral inhibition of D1R-expressing neurons. In combination with reduced DA release, there might be a decline in D1R signalling, leading to an impairment in reversal learning performance by blunting learning from negative feedback.

By contrast with the above findings, intra-NAcS administration of quinpirole improved reversal learning by enhancing learning from positive feedback. The lack of effect of ZM-241385 into the NAcS suggests that quinpirole was affecting pre-synaptic D2R on mesolimbic DA neurons projecting to the NAcS, rather than post-synaptic D2R on MSNs, originally thought to be implicated in learning from negative feedback (Frank et al., 2004). As opposed to post-synaptic receptors, pre-synaptic D2R provide feedback inhibition and regulate DA neuron activity and release (Ford, 2014), hence the improvement could be related to a reduction in DA release into the NAcS (Fig. 4.7 for visual interpretation). Quinpirole-induced effects on DA release could be integrated by both the direct and indirect pathway, which the present results suggest to enhance reversal-learning performance by increasing positive approach behaviour. In support of both pathways being involved in controlling flexible behaviour, I observed that D1R antagonism in the NAcS improved performance in a serial reversal learning task by decreasing the number of errors to reach criterion in rats (Chapter 3).

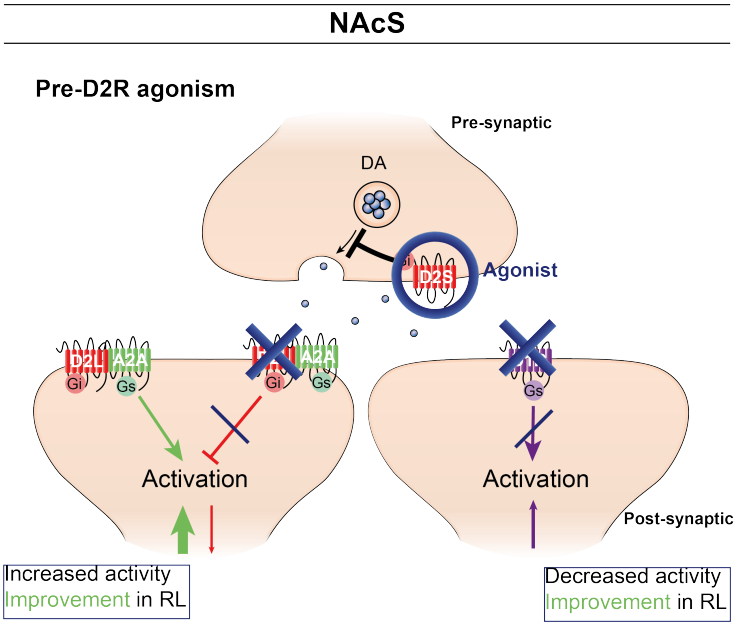


Fig. 4.7 Representation of the potential effect of D2R agonism in NAcS MSNs. Since the post-synaptic D2R-probe agent, ZM-241385, did not replicate the effects of quinpirole, quinpirole must be acting via pre-synaptic D2R. Agonising these receptors inhibits DA release from pre-synaptic neurons, resulting in lower DA levels in the synaptic cleft. A decrease in DA availability, leads to decreased D1R and D2R activity, blocking their respective cell activation and inhibition of the activation. In both cases, this results in enhanced pathway signalling and improvement in reversal learning performance by promoting learning from positive feedback.

Finding an improvement following local quinpirole administration was at odds with my reasoning based on the systemic effects presented in this chapter and previous research. For example, it has been reported that administration of L-DOPA, which increases DA concentration in the synaptic cleft, disrupts reversal learning performance by altering NAc activity in patients with Parkinson's disease (Cools et al., 2007). However, other studies are in line with the observed improvement. In patients with OCD, low levels of DA D2R binding have also been related to poor levels in reversal-learning performance (Denys et al., 2004). Drug abuse disorder patients, who often suffer from cognitive inflexibility (Ersche et al., 2011), show decreased responsiveness to DA (Volkow et al., 2009), suggesting that a boost in DA is necessary to reach normal levels of performance. In rats, pharmacological inactivation of the NAcS impaired performance in a probabilistic reversal learning task (Dalton et al., 2014), as did lesions of the projections innervating this region (Groman et al., 2019), which indicates that the NAc is necessary for optimal reversal learning. Moreover, DA acting in the NAc enhanced conditioned reinforcement (the process by which a stimulus previously associated with reward comes to reinforce instrumental behaviour (Mackintosh, 1974)). Thus, intra-NAc administration of amphetamine, which increases extracellular DA levels, selectively potentiated instrumental responses for a conditioned reinforcer (Taylor and Robbins, 1984), and this effect was blocked by depletion of NAc DA using 6-hydroxydopamine (6-OHDA) (Taylor and Robbins, 1986).

The enhancement in reversal learning following administration of the A2AR antagonist is in line with the enhancement of cognitive flexibility, in both attentional set-shifting and reversal learning, after knocking down A2AR in the NAc (Zhou et al., 2019). Notably, in the present task, whereas quinpirole slowed reward collection latencies, the A2AR antagonist induced the same beneficial effect without slowing motor activity or developing sensitization in later stages of the testing period. The therapeutic implication of this novel finding is discussed in Chapter 7.

The NAcC and NAcS are broadly hypothesised on the basis of several lines of evidence to play opponent roles in modulating behaviour (Dalley and Robbins, 2017). Taken together, the present results from local infusions of quinpirole suggest that these two regions improve reversal learning but *via* two different mechanisms: enhanced learning from negative feedback *via* post-synaptic receptors in the NAcC, and enhanced positive feedback *via* pre-synaptic D2R, potentially resulting from reduced DA release in the NAcS. This suggests that the DA system in the ventral striatum is working in complementary ways in the NAcS *versus* the NAcC, as we also found for ventral *versus* dorsal striatum (Sala-Bayo et al., 2020). The

impairment observed following D2R agonism in the NAcC at higher doses of quinpirole indicates the presence of a complex and dynamic mechanism depending on the level of D2R occupancy, potentially related to the inverted U-shaped function of DA suggested by Yerkes-Dodson principle (Yerkes and Dodson, 1908). Intra-striatal quinpirole has also been shown to improve reversal learning in marmosets with deficits at higher doses (Horst et al., 2019), as well as PET imaging in humans revealed a correlation between striatal DA release and reversal learning performance (Clatworthy et al., 2009).

In addition, while the systemic study was followed by local infusions with the aim of studying the role of striatopallidal or mesoaccumbal MSNs, we cannot exclude the potential interference of D2R in other loci, such as on cortical projections to the striatum or striatal interneurons. Around 95% of neurons within the NAc are MSNs, 5% are interneurons (Castro and Bruchas, 2019), from which 80% express D2R. Selective optogenetics activation of these interneurons enhances phasic DA release in the NAc (Cachope et al., 2012). Furthermore, activation of D2R by tonic DA release selectively inhibits inputs from the PFC, suppressing PFC-dependent set-shifting responses and response inhibition (Goto and Grace, 2005), which may affect performance on the VPVD task. While Frank's model of basal ganglia function (Frank et al., 2004) did not predict the behavioural results from targeting the ventral striatum, it did predict the systemic results. Thus, the systemic quinpirole effect might result from the interaction of different brain regions, rather than having a localised focus, or from the interaction of these regions with a main modulatory structure, such as the dorsal striatum – where direct and indirect pathways expressing D1R and D2R, respectively, are more clearly dissociated –, and this might be working in communication with the ventral striatum to regulate reversal learning.

Interpretation of quinpirole-induced effects on reward collection latencies

Expanding on the local latencies discussed in the previous section, relatively high doses of systemic quinpirole slowed overall latencies to collect earned food pellets. While latencies to respond were not affected, from the dose of 0.025 mg/kg upwards latencies to collect the reward were prolonged relative to the vehicle control group. Altered latencies to collect the reward while responding times are unaffected indicate possible changes in Pavlovian approach motivation. Increased latencies to collect the pellet could indicate decreased motivation for the reward. However, if this were the case, we would expect to find an effect on positive-probe trials, which are the trials measuring approach to rewarded stimulus. The

lack of effects on such trials is at odds with altered motivation and indicates that these effects on performance may be due to cognitive mechanisms.

In contrast, the speed to respond to the stimuli was not affected overall by quinpirole, when collapsed throughout all sessions on trials, which suggests that the drug did not affect the control of choice speed. However, when analysed depending on the type of trial, I observed an increased time to respond regardless of dose on negative-probe trials when animals made an error. In these trials, animals had to dissociate between the negative (A-) and the probe ($C_{50/50}$) stimulus, and it was in those same trials that higher doses of quinpirole impaired performance leading to a reduction on optimal choices. This indicates that animals required more time to make a decision, suggesting that the stimuli processing on those trials was cognitively challenging, but further research would be needed to confirm this hypothesis, as collection latencies were also affected in the same direction.

While these results extend our understanding of the role of D2R and the NAcC and NAcS in reversal learning, a more selective approach is needed to dissociate the involvement of pre- and post-synaptic D2R. Chapter 5 aims to provide evidence for this dissociation by using a systemic A2AR antagonist. In addition, a direct link between modulation of behaviour, changes in DA dynamics, and learning from positive and negative feedback is yet to be established. Chapter 6 seeks to establish this link by artificially stimulating striatal inputs with in-vivo optogenetics during specific times in reversal learning when feedback is processed to guide behaviour.

4.6 Conclusions

This study demonstrated dissociable effects and mechanisms of systemic and striatal D2R activation with quinpirole, and striatal A2AR antagonism with ZM-241385 into the NAcC and NAcS in a visual reversal-learning task in rats. Using the recently developed VPVD task, I observed that high doses of quinpirole impair reversal-learning performance by blocking learning from negative feedback, as predicted by Frank's model of the basal ganglia (Frank et al., 2004). In contrast, I found that whereas low doses of quinpirole into the NAcC improved performance in reversal learning by enhancing learning from negative feedback, quinpirole into the NAcS improved performance by selectively enhancing learning from positive feedback. Administration of KW-241385 suggested that these effects were

modulated by pre-synaptic D2R in the NAcS, potentially leading to reduced DA release, but by post-synaptic D2R in the NAcC. In addition, higher doses of quinpirole into the NAcC impaired reversal learning by blunting learning from negative feedback, as observed following systemic administration of quinpirole.

Chapter 5

Effects of adenosine 2A and dopamine D2 receptor agents on spatial probabilistic reversal learning

5.1 Introduction

Convergent evidence implicates striatal DA as an important neuromodulator of reversal learning (Clarke et al., 2008; McAlonan and Brown, 2003). Electrophysiological experiments in animals have shown that DA signalling corresponds with RPE coding whereby unexpected rewards produce a phasic activation of DA neuronal activity, while unexpected omission of rewards results in a dip in DA activity (Schultz et al., 1997).

In humans, D2R have been widely implicated in reversal learning (Clatworthy et al., 2009; den Ouden et al., 2013; Mehta et al., 2001). Imaging experiments revealed that D2R radioligand binding correlates with learning from negative feedback (Cox et al., 2015). In vervet monkeys and rats, systemic blockade or agonism of D2R impairs reversal learning (Boulougouris et al., 2009; Lee et al., 2007), highlighting an inverted U-shaped function of striatal DA involvement (Horst et al., 2019), while D2R knockout mice show deficiencies in initial visual discrimination and in reversal learning (Kruzich and Grandy, 2004).

Understanding and dissecting the role of DA signalling is challenging due to the expression of D2R in both pre- and post-synaptic striatal MSNs (De Mei et al., 2009; Delle Donne et al., 1996). Both have been shown to regulate DA-evoked release, but only pre-synaptic D2R, also called autoreceptors, regulate DA re-uptake and release (Anzalone et al., 2012). Our work with systemic quinpirole in the VPVD task learning from positive or negative feedback suggested that the D2R agonist quinpirole modulated behaviour *via* post-synaptic D2R (Alsiö et al., 2019), but it is difficult to conclude if the results were modulated by pre- or post-synaptic D2R.

Apart from being expressed in pre- and post-synaptic striatal neurons, D2Rs are also expressed in striatal interneurons (Delle Donne et al., 1996), which play an important role in associative and motor learning processes (Surmeier et al., 2007; Wang et al., 2006). However, a key anatomical difference between interneurons, mesolimbic/nigrostriatal and striatopallidal neurons is that only the latter selectively express A2AR. A2AR are expressed as D2R heterodimers in higher density in striatopallidal neurons than any other cells in the central nervous system (Gerfen, 2004; Schiffmann et al., 2007). Thus, A2AR appear to be specific marker of striatopallidal neurons. A2AR are Gs-coupled receptors and their antagonism has been suggested to enhance the efficacy of DA bound to striatal D2R.

Although selective serotonin reuptake inhibitors (SSRI) are the choice of preference for OCD patients, some subjects show treatment resistance; highlighting that novel neurological disorder (like OCD) treatments are an unmet need that remain challenging (Pallanti et al., 2004). In OCD-like behavior in rodents induced by repetitive administration of quinpirole, blocking A2AR with the antagonist Istradefylline® rescued impaired behaviours that were either responsive or not to the current SSRI treatment, and it improved cognitive flexibility (Asaoka et al., 2019). In this line, epidemiological studies report that caffeine, a non-selective adenosine receptor antagonist, has pro-cognitive and anti-depressant properties, thus has been suggested as self-medication for depressed patients (Leibenluft et al., 1993). Caffeine's action at A2AR has been proposed as an alternative preventive or therapeutic strategy for parkinsonian symptoms (Prediger, 2010). In addition, selective A2AR antagonists are being tested in clinical trials for disorders related to dopaminergic dysfunction such as Parkinson's disease, and positive results suggest they can be used as auxiliary therapies (Hung and Schwarzschild, 2014).

Thus, targeting A2AR in a reversal learning context might contribute to dissociating the receptors by which quinpirole induces its effect on cognitive flexibility, and provide

preclinical evidence for a potential clinical treatment for neuropsychiatric disorders associated with DA dysfunction and/or cognitive inflexibility.

5.2 Aims, approaches, and hypotheses

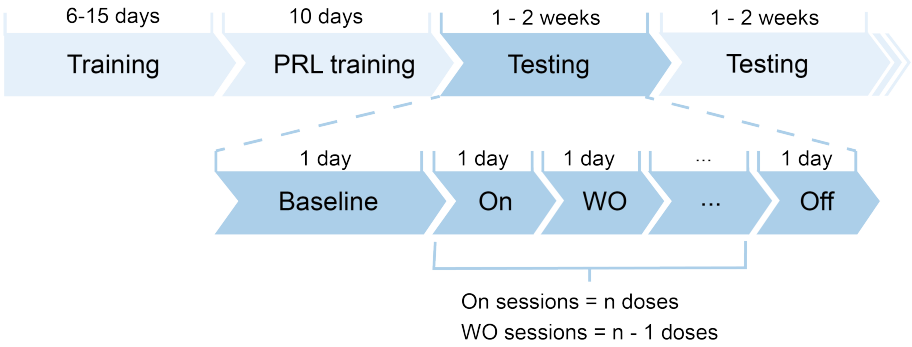
The overarching aim of the experiments described in this chapter was to investigate if the effects of DA agents on cognitive flexibility. Specifically, I sought to research whether the impairments observed in Chapter 4 following systemic quinpirole administration were modulated by striatal post-synaptic D2R, rather than pre-synaptic or interneuron D2Rs. A2ARs form heteromers with D2R in striatopallidal neurons, have been reported to influence DA function, and stimulate the formation of cAMP, whereas D2Rs decrease this cellular cascade. Thus, it was hypothesised that antagonising A2AR would induce a similar effect to agonising D2R (i.e. inhibiting cAMP formation *vs* activating the inhibition of cAMP formation, respectively), if D2R agonism was acting *via* post-synaptic receptors.

This hypothesis was tested by systemically administering the D2R agonist, quinpirole, and the A2AR antagonist, KW-6002 (Istradefylline® being the clinical name), while rats performed in a spatial PRL task. It was predicted that both quinpirole and KW-6002 would impair reversal-learning performance. In addition, it was hypothesised that, the D2R agonist, raclopride, and KW-6002 would counteract each other so their combined effect would not differ from vehicle control.

5.3 Material and methods

All experiments were performed at the CNS Department, in Boehringer Ingelheim GmbH, Germany.

A) Timeline



B) Drugs

Drug	Doses (mg/kg)	Route	Volume	Time pre-treatment	Vehicle
Quinpirole	0 0.025 0.1 0.25	i.p.	1 ml/kg	30 min	Saline 0.9%
Raclopride	0 0.01 0.03 0.1 *0.3*	i.p.	1 ml/kg	20 min	Saline 0.9%
KW-6002	0 0.3 1 3 10	p.o.	5 ml/kg	30 min	1% Methylcellulose 400 cP

C) Cohorts and administered doses

Cohort	Drugs				
	Quinpirole	Raclopride	KW-6002	Quinpirole + Raclopride	KW-6002 + Raclopride
1	n = 16 **Doses:** 0 0.025 0.1 0.25			n = 15 **Doses:** Q0_R0 Q0.1_R0 Q0_R0.3 Q0.1_R0.3	
2			n = 16 **Doses:** 0 0.3 1 3 10		n = 16 **Doses:** KW0_R0 KW3_R0 KW0_R0.3 KW3_R0.3
3		n = 16 **Doses:** 0 0.1 0.03 0.1			

Fig. 5.1 Figure caption on following page

Fig. 5.1 Overview of the drugs used and experimental design. A) Experimental timeline of the behavioural procedure, including training stages to respond to the operant boxes, training in the PRL task, and testing. During testing, for each LSQ, animals underwent a baseline session with administration of vehicle, followed the next day by a session on drug. The day after, animals ran the task off-drug as a washout day (WO). The sequence of 'On-WO' continued until the LSQ was completed i.e. n sessions of 'On' and n-1 sessions of 'WO', where n is the total number of doses included in the LSQ. After this, animals rested for at least one day, as an 'Off' day, before starting the following LSQ, if needed. Time indicated above each cell indicates expected time required to complete the stage. B) Drugs used, including doses, route (i.p.: intraperitoneal; p.o.: per os, oral), volume, time of administration before testing in the operant boxes, and vehicle. Note that the dose of raclopride 0.3 mg/kg was exclusively used when administered in combination with another drug within the LSQ, not when originally tested separately. C) Cohorts size and doses per cohort. Doses were administered following a LSQ design, where each doses was tested during one behavioural session. When the cohort experienced more than one LSQ, the LSQ containing only one drug was tested before that combining multiple drugs. Where n = 15, one rat was excluded due to seizures. Doses are presented as mg/kg.

5.3.1 Subjects

See Chapter 2, section 2.1, for details on housing and ethical approval. A total of 65 male Lister-Hooded rats (Charles River GmbH & Co, Germany) was housed in groups of four under humidity- and temperature-controlled conditions and a 12:12-h light-dark cycle (lights off at 0730 h).

5.3.2 Drugs

The D2R antagonist s(-)raclopride(+)-tartrate salt (Sigma-Aldrich, Dorset, UK) and the D2R agonist (-)-quinpirole hydrochloride (Sigma-Aldrich, Dorset, UK) were dissolved in physiological saline (0.9%). Aliquots were stored at -20°C for a maximum of 1 week in the quantities required for each testing day. The A2A-R antagonist KW-6002 (Istradefylline®, Tocris Bioscience, Germany) was suspended in 1% methylcellulose 400 cP. Doses were prepared daily before testing (Fig. 5.1B). In this chapter, I used KW-6002 instead of ZM-241385 because administration was systemic as opposed to intracerebral (Chapter 4). Systemic ZM-241385 induces peripheral effects before crossing the blood brain barrier (Coney and Marshall, 1998).

5.3.3 Behavioural procedures

Apparatus

The behavioural apparatus consisted of eight lever-pressing operant chambers (Med Associates, Georgia, VT, USA). See Chapter 2, section 2.2.2, for further details on the apparatus.

Training

Rats were initially trained following the standard training procedure for spatial reversal learning. See Chapter 2, section 2.3.2, for details on behavioural training.

PRL task

See Chapter 2, section 2.3.2, for details on the PRL task.

5.3.4 Drug administration and behavioural testing

Fig. 5.1A shows the experimental timeline of the behavioural procedures. After reaching criterion for testing, rats underwent a baseline session following the administration of vehicle (habituation). On the following day, testing started. For all experiments, drugs were administered according to a within-subject LSQ design, fully randomised based on baseline performance. Drug administration was conducted every 48 h to allow for drug wash-out. In between on-drug sessions, an off-drug session was run to maintain performance stability. Raclopride was administered i.p. at doses of 0, 0.01, 0.03, 0.1 and 0.3 mg/kg, 20 min prior to testing. Quinpirole was administered i.p. at 0, 0.025, 0.1 and 0.25 mg/kg, 30 min before the behavioural task. KW-6002 was administered orally, *via* gavage, at 0.3, 1, 3 and 10 mg/kg, in a volume of 5 ml/kg, 30 min before the task. Both i.p. drugs were administered in a volume of 1 ml/kg, whereas the oral drug was administered in a volume of 5 ml/kg (Fig. 5.1B). Suspensions (i.e. KW-6002 doses) were continually stirring prior to administration to avoid the compound settling to the bottom of the vial. In case of combined administration, animals were injected first with the compound of longer waiting time and returned to the cage until

the second administration was needed. Combinations consisted of quinpirole with raclopride, and raclopride with KW-6002 (Fig. 5.1C).

5.3.5 Behavioural data analysis

Behavioural performance was quantified with the dependent variable of the number of reversals either per session or per trial. Trials per session, omissions and latencies to collect the reward were also analysed. I additionally investigated the effect of previous feedback on subsequent decisions, namely the probability of repeat choices after reward ("win–stay") or shifting responses after losses ("lose–shift"). I also looked in detail at responding during the reversal stage. I divided the type of trials following the first reversal into perseverative, correct or incorrect trials, as previously reported (Dhawan et al., 2019; Jones and Mishkin, 1972). Following each reversal, trials prior to the first correct response (i.e. trials touching the previous optimal lever, now suboptimal) were classified as perseverative. Responses on the optimal lever were registered as correct, including the first correct response and disregarding the last eight consecutive correct responses that would lead to a new reversal. Responses on the suboptimal lever, subsequent to the first correct trial were classified as incorrect. Trials were corrected by the number of reversals achieved. Trials from the beginning of the task until the first reversal were analysed as discrimination trials and classified as correct if these were directed to the optimal lever (and up to the first of the eight consecutive trials that led to reversal), or incorrect if responses were done to the suboptimal lever.

Statistical tests were performed using RStudio, version 1.2.1335 (RStudio, Inc). Data were subjected to Linear Mixed-Effects Model analysis with the lmer package in R. The model contained one within-subject factor (dose) and one factor (subject) modelled as an intercept to account for individual differences between rats. For trial type, the model contained two within-subject factors (dose; type of trials) and one factor (subject) as the intercept. Normality was checked with both Q-Q plots and the Shapiro test. To analyse win-stay/lose-shift probabilities, a general model with the glmer package in R was used. Latencies were log transformed while the number of reversals was square root transformed to ensure normality. Homogeneity of variance was verified using Levene's test. For repeated-measures analyses, sphericity was checked with Mauchly's test, and the degrees of freedom were corrected using the Greenhouse-Geisser whenever the sphericity assumption was violated. When significant interactions or main effects were found, analysis was followed by *post-hoc*

pairwise comparisons corrected by a Tukey's test if all the conditions were compared, or by a Dunnett's test if compared against the control group.

5.4 Results

5.4.1 Initial discrimination

Fig. 5.2 shows that high doses of both D2R agonist quinpirole and antagonist raclopride impaired performance during initial discrimination. This within session effect manifested as an increased number of trials to reach the criterion, with increased correct and incorrect responses.

When each drug was administered separately, mixed effects model analysis showed a significant effect of quinpirole dose on the number of trials to reach criterion ($F_{3, 45} = 25.168$, $p < 0.001$), and when raclopride and KW-6002 were administered in the same LSQ ($F_{3, 46.548} = 3.372$, $p = 0.017$) leading to a decrease in the number of reversals achieved. *Post-hoc* multiple comparisons revealed an impairment caused by the high doses of quinpirole: 0.1 mg/kg ($p < 0.001$) and 0.25 mg/kg ($p < 0.001$), but not by the low doses (ns, Fig. 5.2A). For raclopride, only the dose of 0.3 mg/kg impaired performance in comparison with the vehicle control group ($p = 0.028$; Fig. 5.2C). No significant effects were observed with KW-6002 ($p = 0.966$; Fig. 5.2E).

With respect to trial type, there was a significant main effect of dose after administration of quinpirole ($F_{2.097, 31.45} = 18.78$, $p < 0.001$). *Post-hoc* comparisons showed that this effect was driven by all three doses, but in different directions. While the lower dose induced a decrease in the number of trials, and the higher doses induced an raised the number (0.1 and 0.25 mg/kg; $p < 0.001$ for both). I also found a main effect of type of trials with raclopride ($F_{1, 14} = 19.16$, $p < 0.001$) and with KW-6002 ($F_{1, 15} = 17.61$, $p < 0.001$). Whereas *post-hoc* analysis revealed a difference between correct and incorrect trials with the dose of 0.01 mg/kg ($p = 0.047$), this dissociation was not found with any dose of KW-6002 (ns; Fig. 5.2F).

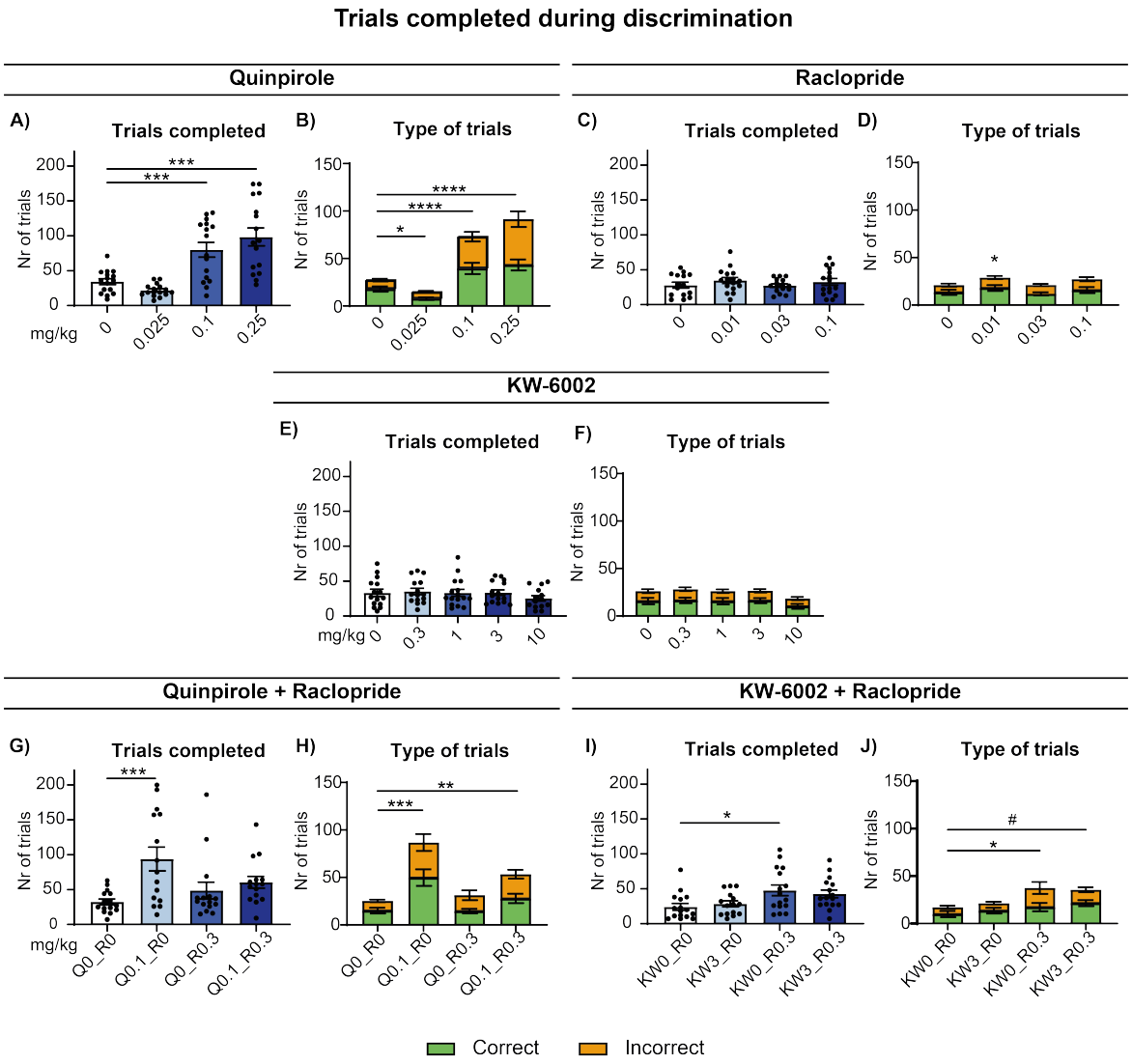


Fig. 5.2 Administration of a higher dose of quinpirole impairs performance to reach criterion to the first reversal. The highest dose of raclopride induces the same effect, except when combined with quinpirole in the same LSQ. For each drug separately: A) Trials completed and B) type of trials following administration of D2R agonist quinpirole; C) Trials completed and D) type of trials following administration of D2R antagonist raclopride; E) Trials completed and F) type of trials following administration of A2AR antagonist KW-6002. For combined drugs: G) Trials completed and H) type of trials following co-administration of quinpirole and raclopride; I) Trials completed and J) type of trials following co-administration of KW-6002 and raclopride. Type of trials include correct trials excluding the eight consecutive trials leading to reversal (bottom; green), and incorrect trials (top; orange). Black line representing significance indicates an overall impairment, including correct and incorrect trials. Data are shown as mean ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ *vs* vehicle treatment.

When drugs were co-administered, the impairment induced by raclopride on discrimination trials was reversed by KW-6002, returning performance to vehicle control levels (dose: $F_{3, 45} = 25.168$, $p < 0.001$; pairwise comparison of combined drugs *vs* vehicle: $p = 1.000$) (Fig. 5.2I). Similarly, the impairment induced by quinpirole was reversed with raclopride (dose: $F_{3, 56} = 5.210$, $p = 0.003$; pairwise comparison of combined drugs *vs* vehicle: $p = 0.215$) (Fig. 5.2G).

In relation to trial type, a significant main effect of dose was found following co-administration of quinpirole and raclopride ($F_{1.827, 25.58} = 8.926$, $p = 0.002$) and of KW-6002 and raclopride ($F_{2.304, 34.56} = 3.215$, $p = 0.046$). *Post-hoc* comparisons revealed these effects were driven by quinpirole ($p < 0.001$), and raclopride did not fully reverse the impairment ($p = 0.002$; in comparison to vehicle control), but attenuated quinpirole effects (quinpirole *vs* quinpirole + raclopride: $p = 0.045$). Following co-administration of raclopride and KW-6002, an overall increase in correct and incorrect trials was observed when raclopride was administered alone ($p = 0.035$), which was reversed when combined with KW-6002 ($p = 0.097$) (Fig. 5.2J).

5.4.2 Reversal learning

Effects of D2R agonism with quinpirole on reversal learning

D2R agonism with quinpirole impaired reversal learning performance at the highest doses tested (0.1 and 0.25 mg/kg), with rats completing fewer reversals than controls per session (Fig. 5.3).

The impairment in performance was observed as a decrease in the number of completed reversals (Fig. 5.3A; $F_{3, 45} = 39.195$, $p < 0.001$), trials per session (Table 5.1; dose: $F_{3, 45.001} = 42.767$, $p < 0.001$) and the ratio of reversals per trials ($F_{3, 45} = 27.203$, $p < 0.001$). Further analysis revealed that this effect was specific to the 0.1 and 0.25 mg/kg dose levels (all $p < 0.001$).

Analysis of win-stay and lose-shift probabilities indicated that quinpirole affected both types of feedback (dose for win-stay: $F_{3, 45} = 41.408$, $p < 0.001$; for lose-shift: $F_{3, 45.002} = 3.722$, $p = 0.018$). *Post-hoc* pairwise comparisons revealed that whereas the dose of 0.1 mg/kg dose decreased win-stay ($p < 0.001$) and increased lose-shift probabilities ($p = 0.029$),

the dose of 0.25 mg/kg decreased win-stay ($p < 0.001$), but left lose-shift probability intact ($p = 0.978$) in comparison to vehicle control. The dose of 0.025 mg/kg did not affect sensitivity to either positive or negative feedback (ns, Fig. 5.3C, D).

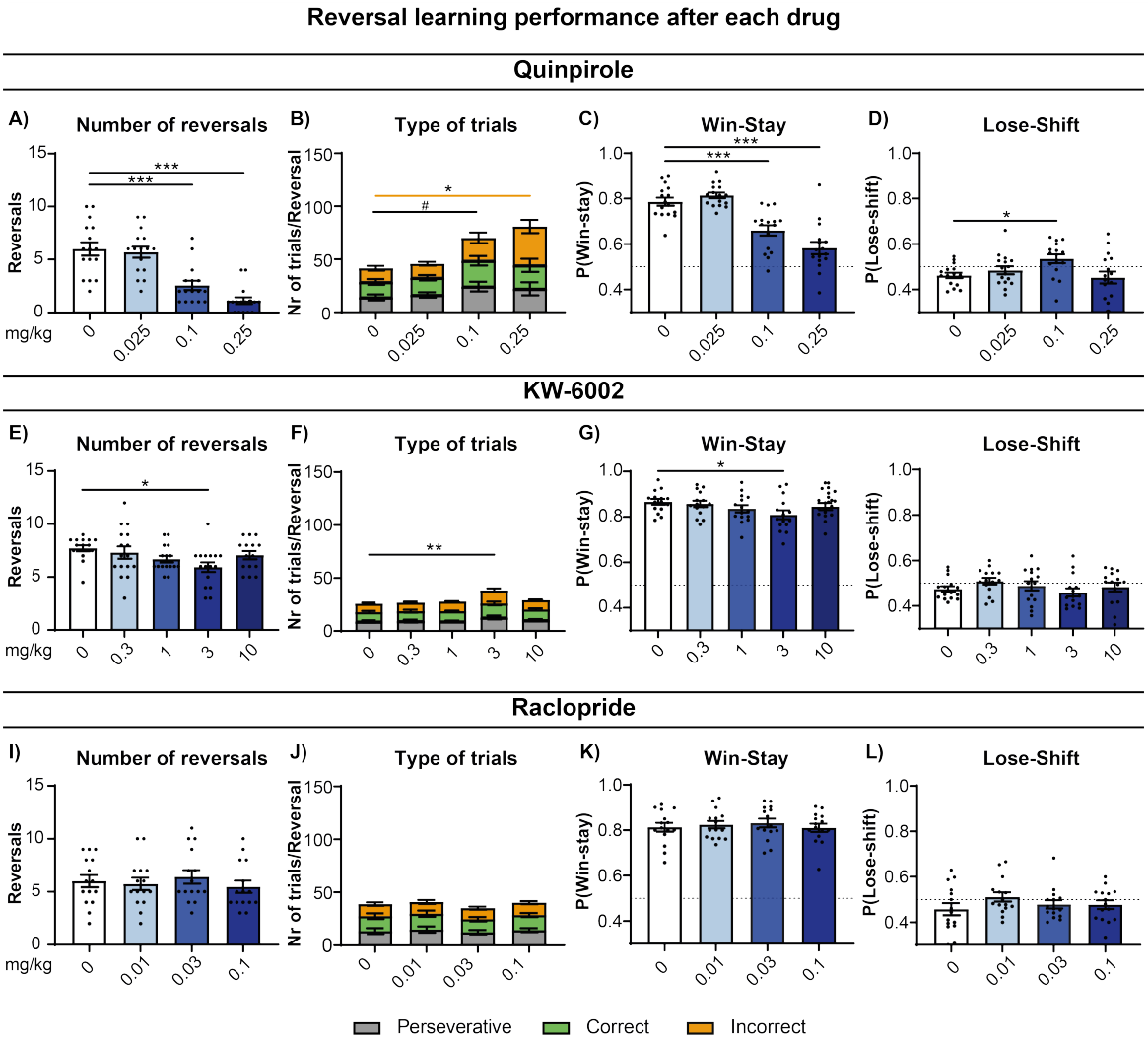


Fig. 5.3 Administration of the three drugs separately impaired reversal-learning performance with relatively high doses. A), E) and I) Number of reversals achieved. B), F), J) Type of trials completed. Perseverative (grey, bottom) trials as touching the new suboptimal (previously optimal) lever following reversal, until touching the new optimal (previously suboptimal) lever; correct trials (green, middle) as touches to the optimal lever until the final error i.e. excluding the eight consecutive responses that lead to reversal; incorrect trials (orange; top) as touches to the suboptimal lever. The colour of the line indicating significance represents what type of trial was significant (i.e. perseverative, correct or incorrect). The line is presented as black if there was an overall significance (i.e. including all type of trials). C), G), K) Win-stay probability. D), H), L) Lose-shift probability. Results are shown as mean ± SEM. $^{\#}$ $p \sim 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ *vs* vehicle treatment.

Group (mg/kg)	Variable			
	Trials	Latency to collect (ms)	Latency to respond (ms)	Omissions
Quinpirole				
0	199.69 ± 0.30	463.11 ± 0.067	350.65 ± 0.01	0.00 ± 0.00
0.025	196.56 ± 1.49	522.10 ± 0.099	469.06 ± 0.02	0.063 ± 0.03
0.1	*181.19 ± 4.47****	587.23 ± 0.110	*954.62 ± 0.06****	*4.25 ± 0.86****
0.25	*155.19 ± 4.10****	677.35 ± 0.289	*1228.07 ± 0.06****	*6.38 ± 0.93****
Raclopride				
0	200.00 ± 0.00	404.03 ± 0.09	400.98 ± 0.09	0.00 ± 0.00
0.01	200.00 ± 0.00	370.12 ± 0.07	367.64 ± 0.07	0.00 ± 0.00
0.03	200.00 ± 0.00	405.96 ± 0.09	405.67 ± 0.08	0.00 ± 0.00
0.1	199.27 ± 0.54	421.64 ± 0.10	421.27 ± 0.11	0.00 ± 0.00
KW-6002				
0	200.00 ± 0.00	455.39 ± 0.05	360.90 ± 0.07	0.00 ± 0.00
0.3	200.00 ± 0.00	453.66 ± 0.05	355.74 ± 0.06	0.00 ± 0.00
1	200.00 ± 0.00	423.78 ± 0.04	349.56 ± 0.07	0.00 ± 0.00
3	200.00 ± 0.00	410.69 ± 0.04	370.56 ± 0.08	0.00 ± 0.00
10	200.00 ± 0.00	393.42 ± 0.04	368.70 ± 0.07	0.00 ± 0.00
Quinpirole + Raclopride				
Q0_R0	199.47 ± 0.52	382.54 ± 0.06	267.92 ± 0.03	0.00 ± 0.00
Q0.1_R0	191.13 ± 2.72	650.50 ± 0.30	597.71 ± 0.23	0.01 ± 0.02
Q0_R0.3	*153.13 ± 11.40****	419.25 ± 0.07	*927.83 ± 0.44****	*0.03 ± 0.05****
Q0.1_R0.3	*154.67 ± 10.06****	510.09 ± 0.08	*784.04 ± 0.35****	0.03 ± 0.04
KW-6002 + Raclopride				
KW0_R0	200.00 ± 0.00	419.11 ± 0.04	337.08 ± 0.05	0.00 ± 0.00
KW3_R0	200.00 ± 0.00	369.19 ± 0.03	336.24 ± 0.06	0.00 ± 0.01
KW0_R0.3	*155.40 ± 7.90****	470.80 ± 0.10	*932.95 ± 0.30****	*0.01 ± 0.03****
KW3_R0.3	199.4 ± 0.58	406.89 ± 0.05	398.77 ± 0.10	0.00 ± 0.00

Table 5.1 Effect of each drug on latencies to collect the reward, latencies to respond and omissions. High doses of quinpirole and raclopride increase latencies to respond and omissions. The joint administration of raclopride and KW-6002 counteracts these effects. The combination of quinpirole with raclopride reverses the increase in omissions, but not in latencies. Data are mean ± SEM. *** $p < 0.001$ *vs* vehicle treatment.

With respect to trial type (Fig. 5.3B), a significant dose × type of trials interaction was found ($F_{1.888,\ 23.60} = 7.998$, $p = 0.003$). Post-hoc pairwise comparisons revealed that this change was driven by the highest dose of quinpirole (0.25 mg/kg) during incorrect trials in each reversal in comparison to vehicle control ($p = 0.009$). A strong trend with the dose of 0.1 mg/kg was also observed in all type of trials (perseverative: $p = 0.050$; correct = 0.063; incorrect: 0.060).

Quinpirole also significantly increased the latency to respond ($F_{3,\ 45} = 51.934$, $p < 0.001$) and omissions ($F_{3,\ 45} = 7.500$, $p < 0.001$) (Table 5.1). Further analysis revealed that this effect were selective for the two highest doses, 0.1 and 0.25 mg/kg (all $p < 0.001$). No changes were observed in these variables after administering quinpirole at 0.025 mg/kg or in latencies to collect the reward with any of the administered doses (ns, Table 5.1).

Effects of A2AR antagonism with KW-6002 on reversal learning

Administration of the A2AR antagonist KW-6002 significantly impaired reversal learning when administered at 3 mg/kg.

With respect to the number of reversals achieved in each session, linear mixed-effects model showed a main effect of dose ($F_{4,\ 74} = 2.704$, $p = 0.037$). Further analysis revealed that this was due to an impairment caused by administration of KW- 6002 at dose 3 mg/kg ($p = 0.011$; Fig. 5.3E). The same effect was observed on reversals per trials ($F_{4,\ 70} = 3.191$, $p = 0.018$). *Post-hoc* comparisons showed that this effect was specific to the dose of 3 mg/kg ($p = 0.010$).

When analysing trial type during each reversal (perseveration, correct, incorrect responses), a main effect of dose was found ($F_{2.913,\ 43.70} = 3.004$, $p = 0.042$), which according to *post-hoc* comparisons was driven by the dose of 3 mg/kg ($p = 0.003$; Fig. 5.3F).

In terms of win-stay probability, there was a close to trend main effect of dose ($F_{4,\ 70} = 1.990$, $p = 0.101$), and planned pairwise comparisons revealed this was driven by the dose of 3 mg/kg ($p = 0.041$). No main effect of dose was detected on completed trials per session, latency to collect the reward, latency to respond, lose-shift probability or omissions (ns; Table 5.1).

Effects of D2R antagonism with raclopride on reversal learning

D2R antagonism with raclopride impaired reversal learning only at the high doses tested. For the raclopride study, the doses of 0, 0.01, 0.03 and 0.1 mg/kg were originally used. These doses proved to have no effect in any of the main analysed variables (Fig. 5.3I-L). A main effect of dose in latency to respond was found ($F_{3, 44.020} = 3.644$, $p = 0.020$), but *post-hoc* pairwise comparisons did not reveal the effect of any dose in comparison to vehicle control, suggesting it was an overall impairment (Table 5.1).

In light of these results, a higher dose of raclopride was tested when combined with the other drugs. The highest dose, 0.3 mg/kg, induced a decrease in reversals per session (dose: $F_{4, 46.379} = 8.703$, $p < 0.001$; Fig. 5.3A, E) and per trials (dose: $F_{4, 46.141} = 4.835$, $p = 0.002$). On type of trials, there was no specific effect of low-dose raclopride or when combined in the same LSQ as quinpirole, but had an overall effect when combined with KW-6002 (see section 5.4.2).

A main effect of dose for omissions was also observed ($F_{4, 46.785} = 9.377$, $p < 0.001$; Table 5.1), latency to respond ($F_{4, 45.257} = 31.228$, $p < 0.001$; Table 1), and win-stay and lose-shift probabilities ($F_{4, 44.597} = 11.518$, $p < 0.001$; $F_{4, 45.520} = 3.181$, $p = 0.022$, respectively; Fig. 5.3G, H). *Post-hoc* analysis revealed that the main effect of Dose was due to the high dose of raclopride: 0.3 mg/kg (all $p < 0.001$, except for lose-shift probability: $p = 0.011$) in all of the significant variables. No effects were observed in latencies to collect the reward (ns; Table 5.1).

Effects of combining D2R antagonism with D2R agonism or A2AR antagonism

After combining the effective dose of the D2R antagonist raclopride (0.3 mg/kg) with the effective dose of the D2R agonist quinpirole (0.1 mg/kg) or of the A2AR antagonist KW-6002 (3 mg/kg), there was a partial recovery of the impairment in reversal learning caused by each drug separately.

The dose of quinpirole 0.1 mg/kg was chosen over 0.25 mg/kg as it induced a similar impairment, but with a reduced impact on latencies or trials, and it tended to affect all type of trials, not only the incorrect (as observed with 0.25 mg/kg). In addition, from previous experiments in our laboratory (Alsiö et al., unpublished), I observed that counteracting the

effects of quinpirole is challenging, so a weaker dose would enable detecting rather small changes than a larger dose. The dose of 3 mg/kg for KW-6002 was used, as it was the only dose found to cause an effect in reversal-learning performance.

With respect to the number of reversals per session, both raclopride and KW-6002 impaired reversal learning (dose: $F_{3,\ 45.250} = 11.817$, $p < 0.001$) (Fig. 5.4). In contrast, when administered together, there was no significant difference compared with the control group ($p = 0.185$; Figure 5.4E). The same conclusion was reached when analysing reversals per trials (dose: $F_{3,\ 45.120} = 6.285$, $p = 0.001$; combine drugs *vs* vehicle: $p = 0.101$), and trials per session (dose: $F_{3,\ 45.120} = 6.285$, $p = 0.001$; combined drugs *vs* vehicles: $p = 0.211$; Table 5.1). Trials per session was originally decreased only by raclopride ($p < 0.001$), not KW-6002 ($p = 1.000$) (Table 5.1). Administering both drugs together also reversed the impairment observed with raclopride alone in omissions and latencies to respond ($p = 1.000$; $p = 0.796$; $p = 0.723$, respectively; Table 5.1). Whereas the reduction in lose-shift probability induced by the D2R antagonist ($F_{3,\ 44.749} = 4.072$, $p = 0.012$; raclopride *vs* vehicle: $p = 0.017$) was prevented by A2AR antagonist ($p = 0.165$; Figure 5.4H), the decrease in win-stay remained significant ($F_{3,\ 44.415} = 14.431$, $p < 0.001$; raclopride *vs* vehicle: $p < 0.001$; combined drugs *vs* vehicle: $p = 0.001$) (Fig. 5.4G, H).

In contrast, when quinpirole and raclopride were co-administered, the effects of each drug were not neutralised by the other drug in any of the analysed dependent variables, hence animals showed impaired performance as when quinpirole or raclopride were administered separately (Fig. 5.4A-D; Table 5.1). This was except for omissions, which did not differ from vehicle control (ns; Table 5.1). In this LSQ, there was no effect in lose-shift probabilities (ns; Fig. 5.4D).

In summary, the results from this chapter indicate that systemic quinpirole impairs cognitive flexibility *via* striatal post-synaptic D2R, as suggested by the replication of its effects by the A2AR antagonist, KW-6002. KW-6002 not only replicated the cognitive effects of quinpirole, but also counteracted the impairment observed after antagonising D2R with raclopride.

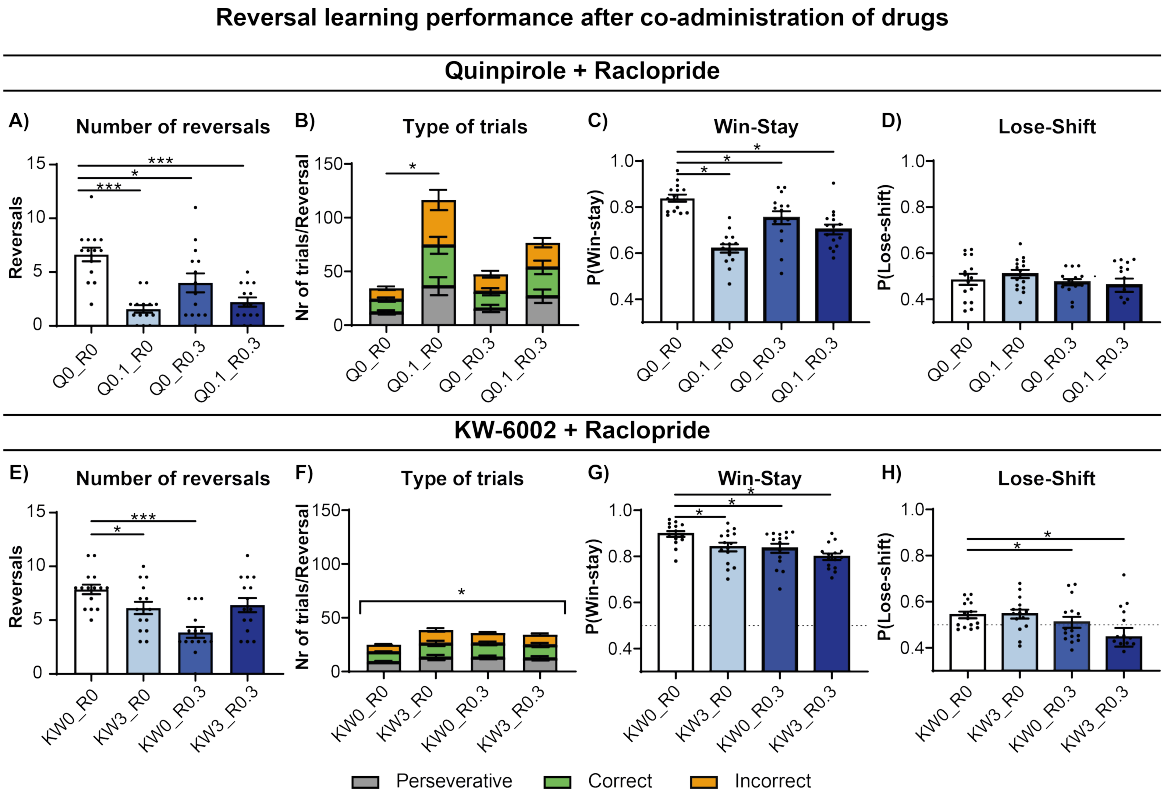


Fig. 5.4 Raclopride and KW-6002, not quinpirole, counteracted their effects when administered together. When drugs are co-administered, doses are represented as the first letters of the drug and followed by the dose in mg/kg i.e. Q: quinpirole, R: raclopride, KW: KW-6002. A), E) Number of reversals achieved. B), F) Type of trials completed. Perseverative (grey, bottom) trials as touching the new suboptimal (previously optimal) lever following reversal, until touching the new optimal (previously suboptimal) lever; correct trials (green, middle) as touches to the optimal lever until the final error i.e. excluding the eight consecutive responses that lead to reversal; incorrect trials (orange; top) as touches to the suboptimal lever. The line for significance is presented as black since there was an overall significance (i.e. including all type of trials); Square bracket line represents significance from main effect of dose, not interaction, and overall impairment. C), G) Win-stay probability. D), H) Lose-shift probability. Results are shown as mean ± SEM. * $p < 0.05$, *** $p < 0.001$ *vs* vehicle treatment.

5.5 Discussion

This chapter demonstrated that stimulation or inhibition of D2R *via* administration of the D2R agonist quinpirole or the D2R antagonist raclopride impaired spatial probabilistic reversal learning performance by increasing the number of trials needed to achieve a reversal. The A2AR antagonist KW-6002 had similar impairing effects, which were blocked by raclopride.

Impairing effects of D2R agonism and antagonism on reversal learning

Quinpirole and raclopride severely impaired probabilistic reversal learning and its preceding spatial discrimination. These results are supported by earlier observations in our laboratory and in previous literature in which D2R agonism and blockade impaired performance in rats (Boulougouris et al., 2009), non-human primates (Smith et al., 1999) and humans (Mehta et al., 2001). These findings on one hand provide a validation of the task for the future optogenetics experiments (see Chapter 6); and on the other hand extend previous findings by showing that the detrimental effects of DA imbalance (e.g. induced by raclopride) in a probabilistic version of reversal learning are counteracted by A2AR antagonism.

Following previous experiments in the VPVD task (Chapter 4) (Alsiö et al., 2019), I reasoned that the impairing effects of quinpirole were mediated by post-synaptic D2R. This assertion is partly supported by the observation that only the higher doses of quinpirole, which putatively act at both pre-synaptic and post-synaptic striatal D2R, rather than lower doses which potentially act pre-synaptically (Eilam and Szechtman, 1989), impaired reversal learning. To test this hypothesis, I combined the administration of D2R agents with the A2AR antagonist KW-6002, clinically known as Istradefylline®. A2AR are structurally coupled with D2R in striatopallidal neurons i.e. exclusively to post-, not pre-, synaptic striatal D2R (Moreno et al., 2018). Since A2AR are Gs-coupled receptors and their antagonism enhances the efficacy of striatal D2R, it is reasonable to assume that the antagonism would replicate the effects of quinpirole if the latter was acting *via* post-synaptic D2R, but would not if it were *via* the pre-synaptic D2R. Similarly, it would counteract the effects of raclopride if this assumption was correct.

KW-6002 reproduced the effects of quinpirole in reversal learning performance and attenuated those caused by raclopride. When administered with raclopride, KW-6002 partially recovered the impairment induced by raclopride – or raclopride of those caused by KW-6002 – in terms of number of reversals achieved. However, it did not affect the decrease in win-stay probability. In addition, the lowest dose of quinpirole – thought to act selectively *via* pre-synaptic D2R – improved performance in the discrimination phase by decreasing the number of correct and incorrect trials, in opposition to the impairment observed with higher doses. Thus, it can be concluded that the observed deficit in probabilistic reversal learning induced by quinpirole is probably modulated by post-synaptic D2R. However, a more complex system or interaction cannot be excluded, and this interpretation should be

made with caution since both raclopride and quinpirole interact with not only D2R, but also D3R.

D2R agonism with quinpirole decreased win-stay probability

Modulation of D2R expressing striatopallidal neurons during reversal was also supported by my studies with systemic quinpirole by the deficit in learning from negative feedback in the VPVD task (Chapter 4). Seminal work from Frank and colleagues (Frank et al., 2004) suggests that the indirect pathway of the striatum mainly contributes to learning from negative feedback i.e. it reinforces "NoGo" behaviours in the presence of negative outcomes. According to the RPE theory, the 'dip' produced during negative outcomes allows the disinhibition of the indirect pathway and acts as a teaching signal. Thus, increased tonic DA activity produced by quinpirole would be expected to prevent learning from such a dip and therefore learning from negative feedback. This is also supported by the decrease in lose-shift probability in rats (Alsiö et al., 2019) and the relationship between D2R and learning from negative outcomes in humans (Cox et al., 2015). Nonetheless, although I observed a comparable impairment in the general performance, in this case, it was driven by a decrease in win-stay probability, not lose-shift. This drug-induced decrease in win-stay indicates that even after receiving a reward, animals treated with quinpirole shifted to the other lever in the following trial more often than when receiving the saline treatment. Although this finding was unexpected, other studies are in agreement with this finding. For instance, D2R binding in the striatum of vervet monkeys correlates with sensitivity to positive feedback in reversal learning, measured as win-stay probability (Groman et al., 2014, 2011). Systemic administration of quinpirole also reduced win-stay probabilities in rats. A decrease in reversals has not only been related to a reduction in the learning rate from losses, but also to decreased learning from wins (Alsiö et al., 2019). In addition, Verharen and colleagues recently reported decreased win-stay probability in a spatial deterministic reversal task following cocaine and amphetamine pre-treatment (Verharen et al., 2018). These behavioural findings suggest that the observed decline in reversal learning in the present study was due to insensitivity of learning from positive feedback, which contributes to optimal performance in the task. Nevertheless, it is important to note that win-stay and lose-shift probabilities are indirect measures of learning from positive or negative feedback, and so both modes of feedbacks should be integrated to produce either win-stay or lose-shift behaviours.

When quinpirole and raclopride were included in the same LSQ, neither affected win-stay or lose-shift probability when administered separately nor combined. Quinpirole is

well known for having long-term effects. I provided a washout day in between sessions to prevent carry-on effects, but this may not have been long enough to prevent residual drug effects on future performance. Since raclopride interacts with the same receptors, they could have neutralised each other regarding their effect on win-stay probability. However, this neutralisation was not observed on reversals. Each drug individually had a strong impairing effect, potentially making it less likely to be reversed. Another speculative possibility would be that quinpirole and raclopride were primarily acting in different brain subregions that co-operate or compete for the modulation of flexible behaviour. Further research would be needed to confirm this hypothesis.

Importance of DA levels for reversal-learning performance

Administration of quinpirole and raclopride, D2R agonist and antagonist, respectively, lead in both cases to an impairment in reversal learning. This is consistent with opposite effects observed on reversal learning after the up- or down-regulation of striatal DA activity. Thus, systemic amphetamine administration, which raises synaptic DA, induced perseverative behaviour following visual object reversal in marmosets (Ridley et al., 1980), whereas selective depletion of caudate DA disrupted reversal in marmosets (Clarke et al., 2011) and rats (O'Neill and Brown, 2007).

The relationship between DA function and behaviour or cognitive output often takes the shape of an inverted-U-shaped function (Yerkes and Dodson, 1908). Recently, Horst and colleagues (Horst et al., 2019) reported that quinpirole acting in D2R has a tri-phasic behavioural effect, in which low and high doses of quinpirole into the caudate in marmosets impaired behavioural flexibility, whereas intermediate doses led to an improvement in reversal performance. Their study showed causal evidence of an inverted U-shaped DA function in cognitive flexibility in a reversal-learning task. Although this inverted U-shape of DA has received criticism in experimental psychology for its capacity to account for various data sets, it conforms to the results found here on reversal performance following administration of D2R agonism and antagonism: quinpirole and raclopride might adjust the optimal level of striatal DA neurotransmission for performance to detrimental levels.

In this study, quinpirole-induced impairment in reversals were accompanied by an increase in perseverative, correct and incorrect trials per reversal at the dose of 0.1 mg/kg, or incorrect responses at the dose of 0.25 mg/kg in comparison to saline treatment. The potential increase in perseverative errors is in line with a decreased learning rate from losses (i.e. α_{loss}

parameter) or increased "inverse temperature" (i.e. β parameter) observed in our previous research (Alsiö et al., 2019). Reduced α_{loss} indicates that animals learn less from negative feedback on quinpirole compared with vehicle control. Since following reversal trials on the previously rewarded lever are now more likely to be unrewarded, pressing that lever would provide negative feedback. Hence, animals with blunted learning from losses would tend to perseverate. In addition, elevated β indicates higher reinforcement sensitivity i.e. less exploration, suggesting that rats on quinpirole were guided by the expected outcome of their responses. In this case, rats would exploit the lever that was optimal (now suboptimal) more instead of exploring the previous suboptimal lever (now optimal). This could similarly apply to the increase in incorrect trials. In summary, the observed impairment could be explained by either or both decreased α_{loss} or increased β parameters. However, for the dose of 0.1 mg/kg a trend to increase the number of correct trials was also observed, which highlights the importance of animals staying on one lever to complete the eight consecutive responses for reversal. For this task, the number of correct trials does not necessarily reflect the quality of performance in reversal learning. Their consecutiveness becomes more relevant, as reversals would only be achieved after eight correct trials in a row. The lack or weaker increase in perseverative errors following raclopride or KW-6002 during reversals indicates that animals could initially inhibit the previously learned choice, but were challenged in maintaining this new choice pattern after initial selection.

The effects of D2R manipulation on discrimination are controversial. During the initial trials up to the first reversal, here named as discrimination trials, there was a selective impairment with the higher doses of quinpirole, and of raclopride when administered in the LSQ with KW-6002, and not with quinpirole. The increase in trials, including both correct and incorrect trials, indicates that animals were impaired at discriminating between the optimal and suboptimal lever, or at remembering this condition from the previous session, making them struggle to find a pattern or stick to one lever to successfully perform in the task. While Lee and colleagues found that administration of quinpirole to monkeys impaired both acquisition and reversal-learning in a three-choice task (Lee et al., 2007), Boulougouris and colleagues observed no differences on acquisition of a spatial reversal learning discrimination after quinpirole treatment (Boulougouris et al., 2009). These discrepancies could be explained by the difference in tasks e.g. two *vs* three-choice discrimination or visual *vs* spatial or by the definition of discrimination. In our study, this discrimination cannot be accounted as acquisition since animals have experienced these conditions and reversals several times prior to the session under quinpirole treatment.

Localisation of the effects in the striatum

Although this study was based on systemic administration of drugs, the observed effects are potentially driven by altered functioning of the striatum, which has been widely implicated in reversal learning performance *via* D2R. The question remains as to which subregion is involved. D2R availability in the dorsal subregions of the striatum of vervet monkeys has been associated with performance in reversal learning (Groman et al., 2011). Within the dorsal striatum, the DMS has been associated with goal-directed actions and is active in the early-mid phases of reversal learning. Instead, the DLS has been associated with habitual behaviours and becomes engaged at later stages (Brigman et al., 2013). In the ventral striatum, increased DA activity or infusion of quinpirole impaired reversal learning in rats (Haluk and Floresco, 2009). Inactivation of the NAcS, but not the NAcC, impaired probabilistic reversal learning performance (Dalton et al., 2014). According to our previous research (see Chapter 3) (Sala-Bayo et al., 2020), whereas D2R antagonism improved early stages of serial reversal learning when administered into the NAcC (not NAcS), it impaired performance in mid stages when infused in the DMS and induced an overall impairment when infused into the DLS. Verharen and colleagues (Verharen et al., 2019) observed that D2R agonism in the NAc, not DMS or DLS, impaired performance in rats in a similar PRL task to the one used in our study. Therefore, the observed effects of quinpirole in the present Chapter may have been mediated within the ventral striatum, with a larger influence of the NAcS; and the effects of raclopride in the NAcC, but most likely the DLS than the DMS from the dorsal striatum. Although speculative, this heterogeneity could account for some of the inconsistencies observed in our results, e.g. quinpirole and raclopride inducing a reduction in win-stay probability when administered in different LSQs, but not when both drugs were administered in the same LSQ.

Higher doses of quinpirole and raclopride affect choice performance, but not food collection latencies

Higher doses of quinpirole ($\geq$ 0.1 mg/kg) and of raclopride (0.3 mg/kg) slowed animals' lever pressing (Table 5.1) and their co-administration failed to overcome this effect. Nevertheless, these doses failed to prolong the latency for reward (food) collection. This suggests that the induced impairment did not depend on decreased motivation or reward sensitivity. In addition, lower but not higher doses are speculated to reduce motivation by acting at pre-synaptic D2Rs to inhibit activity in midbrain DA neurons (Alsiö et al., 2019). This

again supports that the conclusion that post-synaptic D2R mediated the impairing effects of quinpirole and raclopride in the present study.

On top of counteracting the raclopride-induced decline in performance, antagonism of A2AR did not alter latencies or trials completed per session. As also shown by previous research, A2AR antagonism is a promising tool to treat neuropsychiatric disorders such as depression, OCD, or Parkinson's disease (Asaoka et al., 2019; Hung and Schwarzschild, 2014; Leibenluft et al., 1993). Potential therapeutic benefits of clinically targeting A2AR are discussed in Chapter 7.

5.6 Conclusions

In the present Chapter, the PRL task was validated at Boehringer Ingelheim and demonstrated that both D2R agonism with quinpirole and D2R antagonism with raclopride impair reversal-learning performance in a spatial probabilistic reversal-learning task. This was also found following A2AR antagonism with KW-6002 (Istradefylline®), which additionally blocked the effects of raclopride, suggesting that D2R modulate reversal learning *via* striatopallidal neurons i.e. *via* post-synaptic D2R, not pre-synaptic D2R. Whereas raclopride and quinpirole also affected the initial discrimination, trials completed throughout the task and latencies to respond to the stimuli, these were not affected by KW-6002. The lack of effects beyond those in cognitive flexibility suggest that A2AR could be a therapeutic target for treating neuropsychiatric disorders relates to impaired cognitive flexibility due to DA deficiency. Hence, the present study expands our understanding of the neural mechanism underlying reversal learning and contributes to the design of clinical approaches for patients with neurological or neuropsychiatric disorders related to DA dysfunction.

Chapter 6

Mesoaccumbal, but not nigrostriatal, projections mediate reversal learning by regulating behaviour after reward omission: an *in-vivo* optogenetics approach

6.1 Introduction

In the brain, the pattern of DA cell activity is thought to form the basis of RPEs, which act as a teaching signal to update the value associated with stimuli and/or actions (see Chapter 1; (O'Doherty, 2011; Rescorla and Wagner, 1972). DA fibres that originate in the VTA and SNc provide dense, topographic innervation to the striatum (Björklund and Dunnett, 2007; Dahlström and Fuxe, 1964; Groenewegen et al., 1999). The VTA projects preferentially to the NAc (mesolimbic system), whereas the SN projects preferentially to the dorsal striatum (nigrostriatal pathway). The direction and magnitude of the DA neuron response within the midbrain regions depends on the degree to which the reward is expected. When an experienced reward is better than predicted (positive RPE), a burst in DA neuronal activity and DA release occurs, thereby signalling a discrepancy between prediction and experience. Conversely, if the reward is worse than predicted (negative RPE), a dip in DA neuronal firing

is observed (Daw, 2009; O'Doherty et al., 2017; Schultz et al., 1997). An overdose of DA in the synaptic cleft might prevent the dip originated during negative RPEs, observed during reversal stages, and could prevent learning from worse outcomes than expected (Frank et al., 2004; Klanker et al., 2017). However, the circuit level mechanism underlying negative RPEs is poorly understood.

Within the ventral striatum, the NAcS is involved in suppressing actions to non-rewarded stimuli and plays a key role in behavioural flexibility and response to changes in the incentive value of stimuli (Aquili et al., 2014; Floresco et al., 2006). Indeed, inactivation of the NAcS in rats impaired flexibility on a spatial PRL task (Dalton et al., 2014) and activating D2R improves reversal learning (Chapter 4), while inactivation of the NAcC did not affect reversal learning performance in rats in a spatial PRL task (Dalton et al., 2014). This manipulation did, however, slow approach responses to the levers. Antagonising D2R in this subregion improved early stages of reversal learning in a deterministic visual discrimination task in rats (Dalton et al., 2014; Sala-Bayo et al., 2020). Verharen and colleagues found that increased dopaminergic activity in the VTA-NAc pathway impeded learning from reward losses or punishment, highlighting a causal link between negative RPEs and reinforcement learning (Verharen et al., 2018).

The DMS modulates reversal learning and complex processes associated with shifting between different strategies (Castañé et al., 2010; Ragozzino et al., 2002). It is critical for both learning and expressing goal-directed actions (Ostlund and Balleine, 2008) by mediating the association of a response and outcome representation (Balleine and O'Doherty, 2010). The DLS has been strongly related to habit learning and the formation of stimulus-response, as well as stimulus-reward associations (Aosaki et al., 1994; Yin and Knowlton, 2006).

Although previous studies using single-unit electrophysiology, Ca^{2+} imaging, and fast-scan voltammetry established a correlative association between phasic changes of DA neuronal activity, RPEs signalling, and value-based learning, little causal evidence exists. Previous pharmacological and genetic studies aiming to shed light on this topic lacked the required accuracy in terms of cellular and temporal resolution (Iordanova et al., 2006; Takahashi et al., 2009). Indeed, genetically-engineered manipulations (e.g. gene knockdown) generally produce long-term compensatory effects that hinder the evaluation of selective processes (Parker et al., 2010). Moreover, chemogenetic approaches with transfected receptors (e.g. DREADDs) do not provide sufficient temporal resolution (Boekhoudt et al., 2018; Verharen et al., 2018). By contrast, *in-vivo* optogenetics is temporary and spatially highly resolved, and

recently has been used to investigate the causal link between RPEs and associative learning (Chang et al., 2015; Steinberg et al., 2013). Steinberg and colleagues established for the first time a causal role for temporally precise DA VTA signalling in cue-reward learning in associative blocking (i.e. impaired association between CS and US due to the presence of a second CS) and extinction procedures. Chang and colleagues showed that brief pauses can replace negative RPEs in a Pavlovian over-expectation task. However, the use of optogenetics to investigate the mechanisms of RPEs is still nascent and much further research is needed, especially in the context of cognitive inflexibility. For example, it could be used to determine whether brief pauses or increases in firing can account for how RPEs drive behaviour in a reversal-learning task. However, it is unclear how changes in DA neuronal firing affect not only overall reversal learning, but also its constituent subprocesses of learning, such as making a decision, inhibiting prepotent tendencies or learning from the presence or omission of reward. In addition, the mechanism whereby VTA DA neurons interact with diverse brain regions and support RPEs has not been clearly specified.

Learning and translating A-O representations into action needed for reversal-learning performance is amenable to analysis by computational modelling. Indeed, reinforcement learning has been suggested as a tractable computational process underlying trial-and-error learning (Rescorla and Wagner, 1972). One of the most prominent models to analyse sequential learning data is the Q-learning model (Daw, 2009). The use of such models has allowed unravelling information about how organisms solve and learn cognitive tasks, which was not accessible with traditional behavioural analysis (Alsiö et al., 2019; Verharen et al., 2020, 2019). This includes understanding at what rate animals learn from rewarded or unrewarded trials, the likelihood of exploiting or exploring the stimuli depending on the learned rewarding properties, and the tendency of animals sticking to one of the stimuli, choosing the same stimulus as in the previous trial regardless of its rewarding properties.

6.2 Aims, approaches, and hypotheses

The overarching aim of this study was to provide a demonstration of the causal link between midbrain neuronal firing dynamics and reversal learning performance. Specifically, I sought to obtain novel evidence to dissociate the effect of positive and negative RPEs during specific behavioural events on reversal-learning performance.

It was hypothesised that increased VTA-NAcS neuron activation timed to coincide with omission of reward (i.e. negative RPE) would interfere with reversal learning inducing an impairment in performance by blocking the natural dip of DA release that encodes the teaching signal. That is, the impairment would arise from an insensitivity to negative feedback, which the computational model would reflect as a decrease in the learning rate from losses. In contrast, it was reasoned that the neuronal activation aligned with reward during reversals (i.e. positive RPE) either would improve performance by enhancing the natural burst of neurochemical release or would remain unaltered due to ceiling effects. It was further reasoned that increased SN-DMS neuron activation could similarly induce an impairment in performance when activation was aligned with the unexpected omission of reward. Alternatively, it is conceivable that the DMS may not be involved in the PRL task since its key role is to regulate goal-directed actions (Balleine et al., 2007), which hypothetically would be required during earlier stages of training rather than following extensive training (i.e. as probed during the testing stages).

To investigate this hypothesis, an *in-vivo* optogenetics approach was used to test whether activation of the VTA to NAcS or the medial SNc to DMS pathways regulate reversal-learning performance. Specifically, the impact of this intervention was determined after (1) the presentation or (2) omission of reward, after spurious (3) presentation or (4) omission of reward, and (5) prior to the choice response in a spatial PRL task (Fig. 6.1A).

Using computational modelling, I further investigated whether animals learn by positive or negative feedback, or whether other strategies are adopted (e.g. exploration *versus* exploitation or sticking to the previous response regardless of reward outcome). Since it was assumed that stimulating midbrain DA neurons projecting to the striatum would blunt dips of DA and impair learning from negative feedback, I hypothesised that the learning rate from reward omissions would decrease relative to controls when the optogenetic stimulation coincided with the lack of reward.

6.3 Material and methods

All experiments were performed at the CNS Department, in Boehringer Ingelheim GmbH, Germany.

6.3.1 Subjects

A total of 76 male Lister-Hooded rats (Charles River GmbH & Co, Germany) was housed in groups of four under humidity- and temperature-controlled conditions and a 12:12-h light-dark cycle (lights off at 0730 h). After surgery, animals were housed in pairs.

6.3.2 Behavioural procedures

Apparatus

Rats were trained in eight lever-pressing operant chambers (Med Associates, Georgia, VT, USA), as described in Chapter 2 (section 2.3.2). Briefly, the chambers were enclosed within sound-attenuating boxes with a fan for ventilation. Each chamber had two retractable levers, with a light above each lever, flanking the food magazine. Rewards (45 mg sucrose pellets) were delivered to the magazine by a pellet dispenser. Access to the chambers were through a hinged sidewall, and optogenetics cables were connected from the rats to the ceiling of the boxes through a ceiling centred hole.

Training

See Chapter 2 (section 2.3.2) for details in training. Briefly, animals were habituated to make a lever press response to obtain food reward across four different stages of increasing difficulty. During these stages, animals were presented to the probabilistic nature of the task (80%/20%) and were forced to sample both levers to avoid developing a side bias.

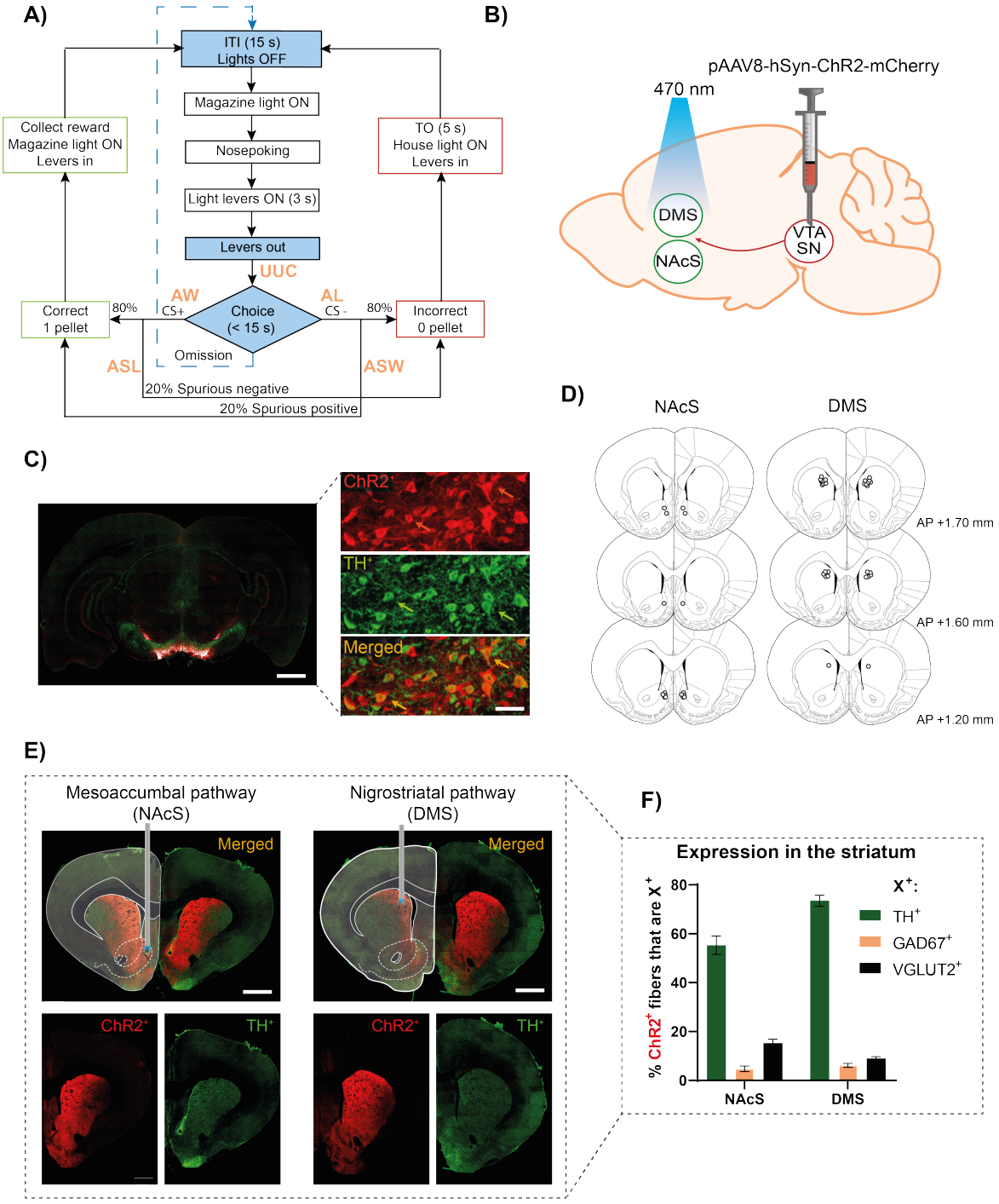


Fig. 6.1 Figure caption on following page

Fig. 6.1 *In-vivo* optogenetic stimulation of the mesoaccumbal and nigrostriatal pathways during a spatial PRL task. A) Flowchart overview of the PRL task. In orange, time-points when the neuronal pathways were optogenetically stimulated after an omission (loss) of reward (AL), after a win (AW), after a spurious loss (ASL) or spurious win (ASW), and up until making a choice (UUC). B) Schematics showing viral vector infusion in the VTA/SN, and fibre optic implantation in the NAcS or DMS for optogenetics stimulation. C) (Left) Coronal section of the VTA stained for TH and showing viral-transfected neurons. Scale bar: 1 mm. (Right top) Detailed expression of virus positive neurons, (Right middle) TH positive neurons (i.e. TH^+), and (right bottom) both channels merged. Scale bar: 50 μm. D) Fibre optic tip placements in the NAcS and DMS. Anteroposterior (AP) coordinates from Bregma. E) Representative histology images showing coronal sections of the striatum and representative fibre optical tip location in the NAcS (Left) and DMS (Right). Expression of viral vector (Bottom left), TH (Bottom right) and both channels merged (Top). Scale bars: 1 mm. F) Percentage co-expression of TH, GAD67 or VGLUT2 in neural fibres expressing the viral vector containing ChR2 (or its inert 'empty' version containing mCherry) in the NAcS or DMS. Data are shown as mean $\pm$ SEM.

PRL task

Behavioural training in the PRL task is described in detail in Chapter 2 (section 2.3.2). Briefly, animals underwent daily sessions consisting of 200 trials or 60 min, whichever came first, including ITI 10 sec, TO 5 sec and LH 10 sec. At the start of each session, one of the two levers was randomly selected to be optimal or suboptimal. A response to the optimal lever delivered a single reward pellet on 80% of the trials, whereas a response to the suboptimal lever gave reward on only 20% of the trials. A failure to press a lever within the limited hold period was noted as an omission. After eight consecutive correct trials (i.e. pressing the optimal lever regardless of this being reinforced or not), the contingency was reversed. Animals were trained to attain at least three reversals over three consecutive sessions before the start of testing.

6.3.3 Stereotaxic surgery

Rats were anaesthetised using isoflurane in oxygen. Anaesthesia was induced with 5% isoflurane and maintained at 2.5%. Animals were ear fixed on a stereotactic frame (KOPF Model 1900, Bilaney Consultants, Germany). An incision was made along the midline of the skin overlying the dorsal skull. The skull surface was cleaned and OptiBond®. All-in-One bone

glue (Kerr, USA) was applied and hardened with UV light for 60 s. Animals were infused bilaterally a total volume of 2400 nL of viral vector per site at a flow rate of 200 nL/min. Control animals received pAAV8-hSyn-mCherry (8.38×10^{12} particles/ml; Boehringer Ingelheim GmbH, Germany) and opsin-expressing pAAV8-hSyn-mChR2-mCherry (7.3×10^{12} particles/ml; Boehringer Ingelheim GmbH, Germany). Animals were divided into VTA-NAcS (n = 30) or SNc-DMS (n = 46) groups. For the first group, the virus was infused into the VTA at AP -5.4 and -6.2, ML $\pm$ 0.6, DV -8.4 and -7.8 (600 nL per infusion), and optical fibres (Doric Lenses, Canada) were implanted in the NAcS at AP +1.5, ML $\pm$ 0.8, DV –7.0. For the second group, the virus was infused into the SN at AP -5.4, ML $\pm$ 0.6; DV -8.1 and -8.0 (600 nL/infusion), and optical fibers (Doric Lenses, Canada) were implanted in the DMS at AP +1.2, ML $\pm$ 2.0, DV –5.3. All coordinates were in mm from bregma and skull. The infuser was left in place for 5 min after each infusion to allow for diffusion. Implants were secured with dental cement, four screws into the skull and Charisma (Kulzer, Germany) hardened with UV light for 20 s to increase the gripping surface for the cement. All surgeries took place before starting behavioural training to allow at least four weeks for viral transfection before testing. After surgery, rats received saline (2 ml once, subcutaneous) and were single housed for the first 3 days of recovery. Training started after allowing at least seven days of recovery.

6.3.4 Behavioural testing

Rats received at least two habituation training sessions with the cables attached prior to testing. A baseline session occurred on the day before testing with the cables attached but without optical stimulation. For each session, the optimal lever was the same as the optimal lever at the end of the previous session. All rats underwent five different optogenetics stimulation conditions (Fig. 6.1B) that were maintained throughout each session and compared against a "light off" (Off) session. These were: (1) "after loss" (AL) when pressing the suboptimal lever, (2) "after win" (AW), when pressing the optimal lever, (3) "after spurious loss" (ASL) when pressing the optimal lever but not receiving the expected reward (20% of the times), (4) "after spurious win" (ASW) when pressing the suboptimal lever but receiving an unexpected reward (20%), and (5) "up until choice" (UUC) from the start of the trial until the choice of a lever. For testing, all conditions were pseudo-randomised according to baseline levels of performance using a LSQ design ((LSQ1) Off – AL - AW; (LSQ2) Off – ASL – ASW) or cross-over design ((LSQ3) Off – UUC, called here LSQ for simplification in below figures

and text/purposes). Testing took place every second day. During days between each test sessions, animals ran with cables attached but with the laser light off to habituate animals to the test apparatus and avoid possible carry-over effects of the light.

6.3.5 Optical stimulation

Mono fibre-optic patch-cord cables (Doric Lenses, Canada) were metal shielded and terminated with an optical fibre of 200 μm of diameter and with a numerical aperture of 0.37 NA. One end of each cable was connected to a PlexBright dual LED commutator via magnetic Blue (λ = 470 nm, max current 200 mA) PlexBright compact LED modules (Plexon, Campden Instruments, UK). A computer running Med PC IV (Med Associates) software, which also recorded responses at both levers and magazine, controlled the optical stimulation. A second computer controlled the behavioural task via transistor-transistor logic (TTL) signals.

Patch-cord cables were covered with 16 cm plastic tubes to prevent animals bending and interfering with the cables. Cables were secured to the rats' implants with a fitted zirconia sleeve (Doric Lenses, Canada) for 1.25 mm diameter ferrule. Intracranial stimulation was achieved with 20 repetitions of 5-ms light pulses (thus 20 Hz), with 2 mW at the tip of each optical fibre, which was checked before and after each photostimulation session. Fig. 6.2 shows the timeline of the experimental design for the optical stimulation. Data from sessions where light output was compromised because of broken or disconnected optical cables were discarded. This criterion led to the exclusion of rats from both groups, leading to the following final numbers per experiment: DMS (LSQ1) n = 18, (LSQ2) n = 15, (LSQ3) n = 26; and NAcS (LSQ1) n = 15, (LSQ2) n = 13, (LSQ3) n = 13. The optogenetics light condition was a between-subject factor and the same animals could be tested in multiple LSQs.

6.3.6 Histological assessment of fibre-optic probe placement and viral vector expression

Following completion of the behavioural procedures, animals were anaesthetised with a lethal dose of pentobarbital (Narcoren, Boehringer Ingelheim GmbH, Germany) and perfused transcardially with 0.01 M phosphate-buffers saline (PBS) followed by 4% paraformaldehyde

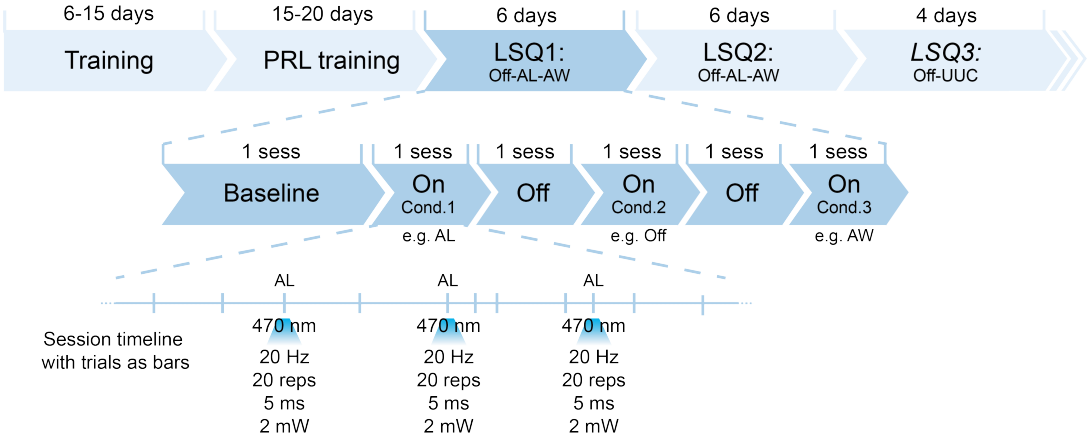


Fig. 6.2 Experimental timeline of the behavioural procedure and optical stimulation. Following pre-training stages to respond to the levers, animals train in the PRL task. In later sessions of this stage, animals are connected to the patch-cord cables without light. Once performance was stable (i.e. $\geq$ 3 reversals in 3 sessions in a row), testing started. Testing consisted on two LSQs and a crossover design, labelled here as LSQ3. Each LSQ started with a baseline session consisting on a PRL session with the cables connected to the rats, without light. The different conditions (e.g. Off-AL-AW) were then tested. Light (λ = 470 nm) was selectively turned on after each trial that presented the tested condition e.g. after not receiving a reward (AL). Intracranial stimulation was induced with 20 repetitions of 5-ms pulses, with 2 mV at the tip of the optical fibres. Testing conditions within each LSQ (or crossover design) were pseudo-randomised. Final group sizes: DMS (LSQ1) n = 18, (LSQ2) n = 15, (LSQ3) n = 26; and NAcS (LSQ1) n = 15, (LSQ2) n = 13, (LSQ3) n = 13. Abbreviations: after loss (AL), after win (AW), after a spurious loss (ASL), after a spurious win (ASW), up until making a choice (UUC), session (sess).

(PFA) administered with a pump flow rate of 8 ml/min. Brains were removed and post-fixed in PFA for 24 h and dehydrated for cryoprotection in 30% sucrose in 0.01 M PBS.

Brains were sectioned coronally at 60 μm using a cryostat (Leica, Germany), collected in PBS containing 25% polyethylene glycol and 25% glycerol, and stored at + 4°C. Free-floating sections were washed in PBS and subsequently blocked and permeabilised in PBS containing 3% normal goat serum (NGS) and 0.3% Triton for 1 h. Sections were incubated overnight with primary antibodies in PBS containing 3% NGS and 0.3% Triton. Since the infused viral vector inherently expressed fluorescence, no antibodies were required. For a first set of slices, TH was detected with the primary antibody anti-TH in rabbit (1:600, EMD Millipore - Merck, USA). After washing in PBS, sections were incubated with secondary antibodies for 2 h (anti-rabbit in goat Alexa-Fluor 488 nm, 1:500, Invitrogen Thermo Fisher Scientific, USA). For a second set of slices, double staining was achieved with the primary

antibodies anti-GAD67 in mouse (1:1000, Invitrogen Thermo Fisher Scientific, USA) and anti-VGLUT2 in rabbit (1:600, Invitrogen Thermo Fisher Scientific, USA). As secondary antibodies, goat anti-mouse (Alexa-Fluor 647 nm, 1:500, Invitrogen Thermo Fisher Scientific, USA) and goat anti-rabbit (Alexa-Fluor 488 nm, 1:500, Invitrogen Thermo Fisher Scientific, USA) were used. After washing in PBS, sections were mounted in distilled water and covered with mounting medium (DAPI, EMD Millipore, USA) and a coverslip. Immunofluorescence sections were checked and digitised using a PerkinElmer Opera Phenix High-Contrast Screening microscope (PerkinElmer, USA). Cell counting was achieved using Columbus software (PerkinElmer, USA).

6.3.7 Computational modelling

Model parameters

To model the computational processes underlying reversal learning performance, several reinforced learning algorithms introduced previously were used (Zhukovsky et al., 2019), including three variants of the Q-learning model (Daw, 2009) defined below, and three parameters: α, β, and κ. The learning rate α determines the degree to which animals learn in response to feedback. In model 3, the learning rate was split into α_{Reward} or α_{R}, and α_{NoReward} or α_{NoR}. The learning rate α_{R} determines how quickly the model adjusts the expected Q-value of a response following the receipt of a reward (positive feedback), while the learning rate α_{NoR} determines how quickly the expected Q-values were adjusted following a non-rewarded response. These expected Q-values were converted into action probabilities using the softmax rule with the inverse temperature parameter β and the choice autocorrelation or "stickiness" parameter κ. In this implementation of the model, high β values result in random exploration of all options, and down weight the contributions of the expected Q-values to the probability of choosing a given action. Low β values result in greater exploitation of the chosen action. In the present reversal task, with 80/20 probabilistic outcomes, a high β value would result in fewer rewards. Finally, the choice autocorrelation parameter κ is a measure of "stickiness", or how likely an animal will perform the same response again regardless of reward outcome. Values of κ close to 1 reflect an agent "sticking" to the previous response while κ values close to -1 reflect choice alternation. In the 80/20 probabilistic reversal task, a moderately high "stickiness" is advantageous as it leads agents to ignore the spurious losses or reward omissions. Model parameters were fitted to each

animal's individual data and compared using analysis of variance (ANOVA).

Model-free Q-learning: model 1

Simple Q-learning is equivalent to Rescorla-Wagner learning (Rescorla and Wagner, 1972) whereby an agent assigns an expected Q-value to each choice available; presently a left or right response (L or R) at each trial t. The expected Q-value is updated on each trial according to the following:

$$Q_{t+1}(c_t) = Q_t(c_t) + \alpha(r - Q_t(c_t)) \tag{6.1}$$

where $0 \leq \alpha \leq 1$ is a learning parameter, $Q_t(c_t)$ is the value of the choice c_t at trial t and r takes the value of 1 if the choice was rewarded and a value of 0 if not. A large α implies faster updating of the expected Q-values of a response after a trial is completed. The probability of making the choice c_t at trial t was calculated using the softmax rule:

$$P(c_t = L|Q_t(L), Q_t(R)) = \frac{exp(Q_t(L)/\beta)}{exp(Q_t(L)/\beta) + exp(Q_t(R)/\beta)} \tag{6.2}$$

whereby larger β values lead to more exploration of the responses with lower Q-values. On the other hand, smaller β values result in exploitation of the response with higher Q-values.

Model-free Q-learning: model 2

Model 2 differed from Model 1 only in including the "stickiness" parameter (κ) in the observational part of the model in addition to β:

$$P(c_t = L|Q_t(L), Q_t(R), L_{t-1}, R_{t-1}) = \frac{exp(Q_t(L)/\beta + \kappa L_{t-1})}{exp(Q_t(L)/\beta + \kappa * L_{t-1}) + exp(Q_t(R)/\beta + \kappa * R_{t-1})} \tag{6.3}$$

Whereby a larger κ results in greater probability of the choice c_t at trial t being the same as the choice c_t at trial $t-1$. The same approach was applied to the right sided choice.

Model-free Q-learning: model 3

Model 2 was extended to include a separate α for learning from rewards and losses, α_{R} and α_{NoR}, depending on whether the animal received a reward on trial t. The decision probability was updated in the same way as in Model 2.

Model fitting and comparison

More details on the model fitting and comparison can be found in (Zhukovsky et al., 2019) and (Daw, 2009). Briefly, the parameters were fitted to maximize the probability of data D (the product of the individual probabilities of making a choice c_t at trial t) by finding the maximum of the probability density function $arg\,(\underset{\theta}{max})\,P(D|M,\theta)$.

$$P(Data\,D|Model\,M,\,parameters\,\theta) = P(D|M,\,\theta)\,=\,\Pi P(c_t|Q_t(L),Q_t(R)) \tag{6.4}$$

Model space was treated as discrete, using the following range of parameters: α_{R} and α_{NoR} [0.001 0.005 0.01 0.015 0.02 0.04 0.06 0.08 0.1 0.15 0.2 0.25 0.3 0.4 0.5 0.6 0.7 0.8 0.85 0.9 0.95 0.99]; $0.005 \leq \beta \leq 5$ with a step size of 0.1 and $-1 \leq \kappa \leq 1$ with a step size of 0.05. The parameter range was chosen based on the a priori expectations regarding α and κ, as well as empirical information about best-fit β parameters. More refined ranges were also tested, without a clear advantage on results accuracy. Model selection was conducted using the Bayesian Information Criterion (BIC) that incorporates the likelihood of data given the model with the best fit parameters $(P(D|M,\hat{\theta}_M))$ and a penalty term $\frac{n}{2}log\,m$:

$$BIC = |log(P(M))| \approx log(P(D|M,\hat{\theta}_M)) - \frac{n}{2}log\,m \tag{6.5}$$

where n = number of free parameters and m = number of observations. We also report a biased measure of model fit, pseudo r^2. Scripts implementing the models were written in Matlab R2016a and can be found in the following link: https://github.com/peterzhukovsky/reversal_learning).

Model validation using simulations

To further assess the validity of the winning model, a set of simulations was used. Specifically, groups of rats (n = 40 rats/group) were simulated with parameter values randomly taken from the distribution of the estimated parameters from each opsin group and light condition in the actual experiment. Then, each simulated rat completed the PRL task in a virtual environment, updating the Q values and the probabilities of choosing left or right depending on the four individual parameters of that particular rat (i.e. α_{R}, α_{NoR}, β and κ for model 3) and a trial-by-trial accumulation of information, including reward probabilities (i.e. 80%/20%) and reversals after eight consecutive responses to the optimal lever.

6.3.8 Behavioural data analysis

Behavioural performance was quantified with the dependent variable of the number of reversals either per session or per trial. Trials per session, omissions, and latencies to collect the reward were also analysed. We additionally investigated the effect of previous feedback on subsequent decisions, namely repeat choices after reward ("win–stay" probability) and shifting responses after losses ("lose–shift" probability).

Statistical tests were performed using RStudio, version 1.2.1335 (RStudio, Inc). Data were subjected to Linear Mixed-Effects Model analysis with the lmer package in R. The model contained a two fixed factors (group, light) and one factor (subject) modelled as an intercept to account for individual differences between rats. Normality was checked with both Q-Q plots and the Shapiro test. To analyse probabilities (e.g. win-stay/lose-shift), I used a general model with the glmer package in R. Latencies were log transformed to ensure normality. Throughout the experiment, there were no significant differences between reversals per session and reversals per trial. However, reversals per trial (or x100 trials) followed a normal distribution after being squared root transformed, whereas reversals per

session was not following a normal distribution regardless of the modification. Furthermore, for purposes of visual presentation, the findings are presented as reversals/100 trials.

Homogeneity of variance was verified using Levene's test. For repeated-measures analyses, sphericity was checked with Mauchly's test, and the degrees of freedom were corrected using the Greenhouse-Geisser whenever the sphericity assumption was violated. When significant interactions were found, analysis was followed by *post-hoc* pairwise comparisons corrected by a Tukey's test if all the conditions were compared, or by Dunnett's test if compared against the control group.

Group sizes were based on the minimum number of animals required to obtain statistically reliable results. These were informed by power analyses with significance = 0.05 and power = 0.8, using expected effect sizes based on our own preliminary data or other laboratories' studies, which resulted in required group sizes of 10 subjects (Cohen, 1988). As the success of obtaining results depended on To account for the risks of transfection, targeting of optic fibres, implants durability and animal behaviour, nine extra rats were included in each group, leading to an initial group size of n = 19. Power calculation with the final group sizes was conducted using the pwr package in R based on previous work by Cohen (1988).

6.4 Results

6.4.1 Power calculation

After excluding animals due to poor performance, lost implants or misplaced viral expression or optical fibres, final group sizes ranged from 7 to 14 for all the conditions. For the NAcS, group sizes varied between 7 and 8 rats per group. For the DMS, group sizes varied between 7 and 14. From a power analysis, power was calculated to vary from 0.60 (n = 7) to 0.93 (n = 14).

6.4.2 Viral expression and fibre optic placement

Figure 6.1C, E, F show that the administration of the opsin-expressing virus into the VTA or SNc resulted in high expression of ChR2. Although the infusion coordinates were different for each region, both regions showed high viral expression.

To identity the neurochemical phenotype of neurons infected, the expression of virus and neuronal markers in the neuronal fibres was measured in the DMS and the NAcS. In the DMS, 73.51 ± 2.26% of transfected neurons were TH positive, 8.99 ± 0.69% were VGLUT2 positive, and 6.17 ± 0.83% were GAD67 positive. In the NAcS, 55.28 ± 3.74% of transfected neurons were TH positive, 15.24 ± 1.56% were VGLUT2 positive, and 4.77 ± 1.13% were GAD67 positive (Fig. 6.1F).

Fig. 6.1D shows fibre optical tip placements in the DMS or the NAcS for those animals that completed the study. Rats with misplaced fibre optic cannula or inappropriate expression were excluded (n = 1 in the NAcS; n = 2 in the DMS).

6.4.3 Behavioural data

Fig. 6.3A indicates that optical stimulation of the VTA-NAcS pathway during reward omission impaired reversal-learning performance by reducing the number of reversals in each session. Activation of this pathway during other trial types did not cause a change in performance (Fig. 6.3B-I). In addition, no significant effects were observed on performance when the DMS was stimulated.

Activation of the VTA-NAcS pathway decreases reversal after reward omission

Fig. 6.3 shows that stimulating the VTS-NAcS pathway impaired performance by reducing the number of reversals attained. For the first set of light conditions (Fig. 6.3A-C), i.e. light on AL or AW, a significant group × condition interaction (group: opsin *vs* control; condition: Off *vs* AL or AW) for the number of reversals relative to the trials per session was present ($F_{2, 26} = 5.985$, $p = 0.007$). There was also a trend in significance for the group × condition interaction in the lose-shift probability, ($F_{2, 26} = 2.500$ $p = 0.100$) win-stay probability ($F_{2, 26}$

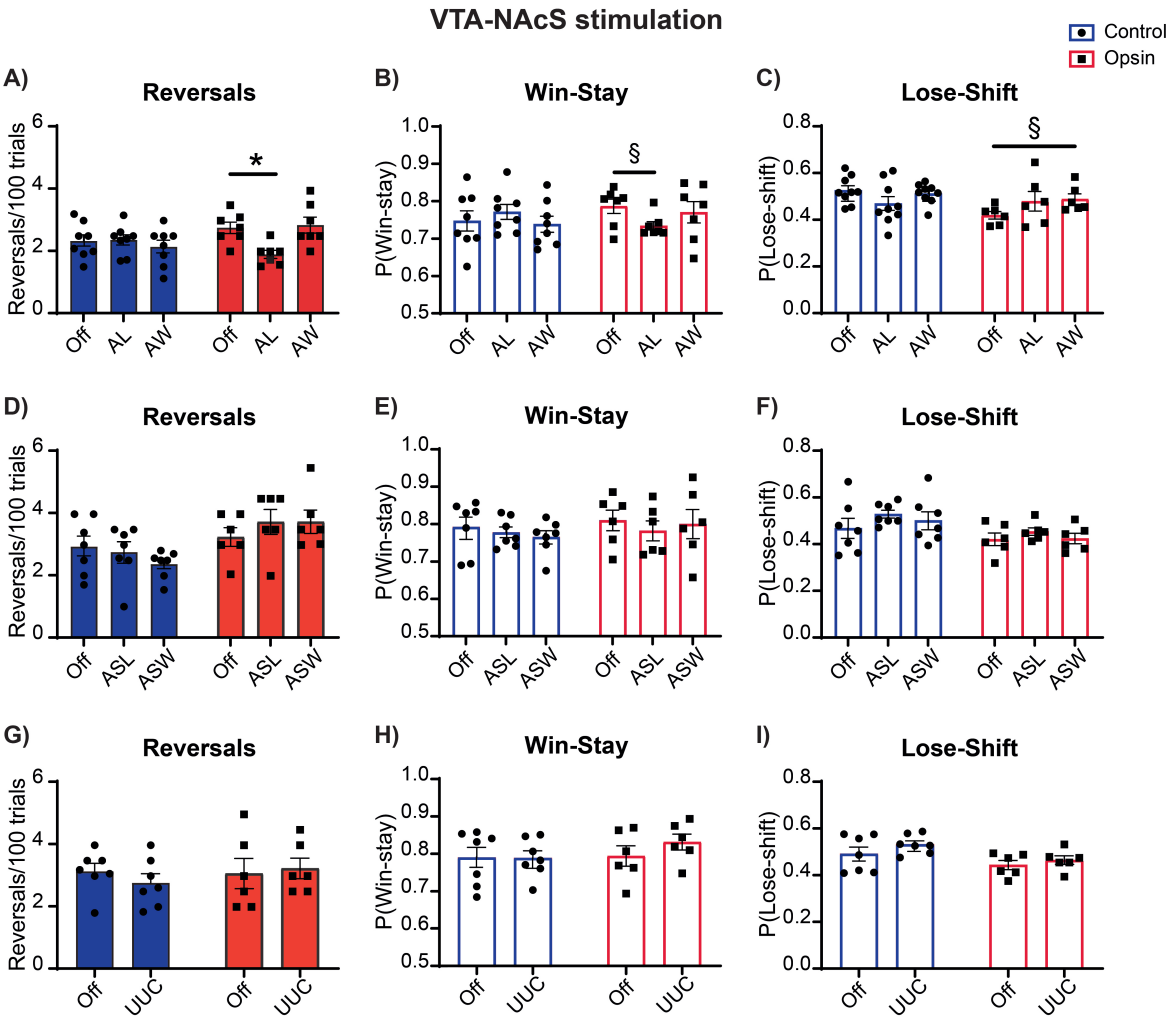


Fig. 6.3 *In-vivo* optogenetic stimulation of the mesoaccumbal pathway impairs reversal learning when selectively activated after a lack of reward. A), D), G) Reversals achieved during each session over 100 trials; B), E), H) win-stay probabilities; and C), F), I) lose-shift probabilities for the different three LSQs to test all light conditions. Light conditions: light off (Off), after lack of reward (AL), after a win (AW), after spurious lack of reward (ASL), after a spurious win (ASW) and up until making a choice (UUC). Data are shown as mean ± SEM. * $p < 0.05$ *vs* light off; § $p < 0.05$ *vs* light off after a trend in group × light condition.

= 2.119, p = 0.131) and in latencies to collect the reward (group × condition: $F_{2, 26} = 2.896$, p = 0.073).

Post-hoc analysis revealed that optogenetics stimulation of the mesolimbic pathway selectively impaired performance by decreasing the number of reversals when the light was on after a lack of reward (p = 0.007, Fig. 6.3A). This was matched by an increase in the latency to collect the reward with the light on after lack of reward (p = 0.012, Table

1). Further analysis showed a decrease in win-stay probability when the light was on after reward omission ($p = 0.039$, Fig. 6.3B). There was also an increase in lose-shift probability specifically when optogenetics stimulation was matched with a win ($p = 0.034$, Fig. 6.3C). There was no difference in terms of trials per session (group $\times$condition: $F_{2, 26} = 0.402$, $p = 0.673$), omissions (group $\times$ condition: $F_{2, 39} = 0.009$, $p = 0.991$) or latencies to respond (group $\times$ condition: $F_{2, 26} = 0.245$, $p = 0.785$).

We observed no behavioural effects with optogenetic activation following the other two sets of light conditions: Off *vs* ASL or ASW, or Off *vs* UUC, in any of the behavioural parameters analysed (Fig. 6.3; Table 6.1).

Region	Condition	Latency to collect (ms)		Latency to respond (ms)		Omissions	
		Control	ChR2	Control	ChR2	Control	ChR2
NAcS							
LSQ1	Off	57.14 ± 18.27	64.74 ± 36.05	45.86 ± 34.79	41.46 ± 30.31	0.000 ± 0.000	0.000 ± 0.000
	AL	56.88 ± 20.96	73.49 ± 55.00	46.50 ± 42.21	43.45 ± 28.69	0.000 ± 0.000	0.000 ± 0.000
	AW	56.32 ± 23.80	67.28 ± 38.30	45.32 ± 30.96	41.80 ± 23.81	0.125 ± 0.117	0.142 ± 0.132
LSQ2	Off	63.46 ± 25.80	72.10 ± 33.67	50.08 ± 44.95	42.22 ± 23.00	0.429 ± 0.187	0.000 ± 0.000
	ASL	66.18 ± 25.47	72.09 ± 40.17	49.39 ± 44.52	41.63 ± 20.88	0.000 ± 0.000	0.000 ± 0.000
	ASW	66.68 ± 25.80	73.40 ± 41.97	52.33 ± 48.76	42.81 ± 25.16	0.142 ± 0.133	0.000 ± 0.000
LSQ3	Off	65.81 ± 23.81	71.35 ± 37.62	49.16 ± 29.52	44.04 ± 26.67	0.000 ± 0.000	0.000 ± 0.000
	UUC	68.86 ± 27.25	76.82 ± 56.25	48.26 ± 34.66	46.17 ± 27.88	0.429 ± 0.187	0.000 ± 0.000
DMS							
LSQ1	Off	52.40 ± 46.29	56.54 ± 34.98	36.63 ± 28.59	44.17 ± 35.29	0.000 ± 0.000	0.222 ± 0.629
	AL	55.85 ± 45.93	60.96 ± 67.67	36.21 ± 31.13	37.14 ± 27.84	0.000 ± 0.000	0.000 ± 0.000
	AW	56.73 ± 48.95	55.77 ± 22.25	35.90 ± 21.27	37.11 ± 22.48	0.125 ± 0.117	0.142 ± 0.132
LSQ2	Off	32.76 ± 42.66	34.71 ± 23.77	23.31 ± 16.57	26.71 ± 24.38	0.000 ± 0.000	0.222 ± 0.629
	ASL	32.62 ± 46.47	33.18 ± 19.30	22.86 ± 14.61	26.18 ± 25.52	0.111 ± 0.105	0.000 ± 0.000
	ASW	32.91 ± 41.63	36.24 ± 30.06	23.76 ± 20.21	26.71 ± 24.38	0.111 ± 0.105	0.111 ± 0.314
LSQ3	Off	63.08 ± 65.02	58.81 ± 18.93	39.84 ± 23.24	45.85 ± 22.96	0.000 ± 0.000	0.000 ± 0.000
	UUC	59.40 ± 60.17	58.74 ± 16.28	40.19 ± 25.06	47.56 ± 31.07	0.000 ± 0.000	0.000 ± 0.000

Table 6.1 Lack of effect of SN-DMS stimulation on latencies to collect food reward, latencies to respond, and omissions. Data are shown as means $\pm$ SEM.

No effect of SNc-DMS stimulation on reversal learning performance

Activation of the nigrostriatal pathway had no effects on reversal learning performance (Fig. 6.4). However, in relation to the number of reversals achieved, there was a strong trend for a main effect of condition following stimulation of the nigrostriatal pathway up until making choice (condition: $F_{2, 24} = 4.244$, $p = 0.050$). We also detected an overall decrease in the win-stay probability (condition: $F_{2, 24} = 5.054$, $p = 0.034$).

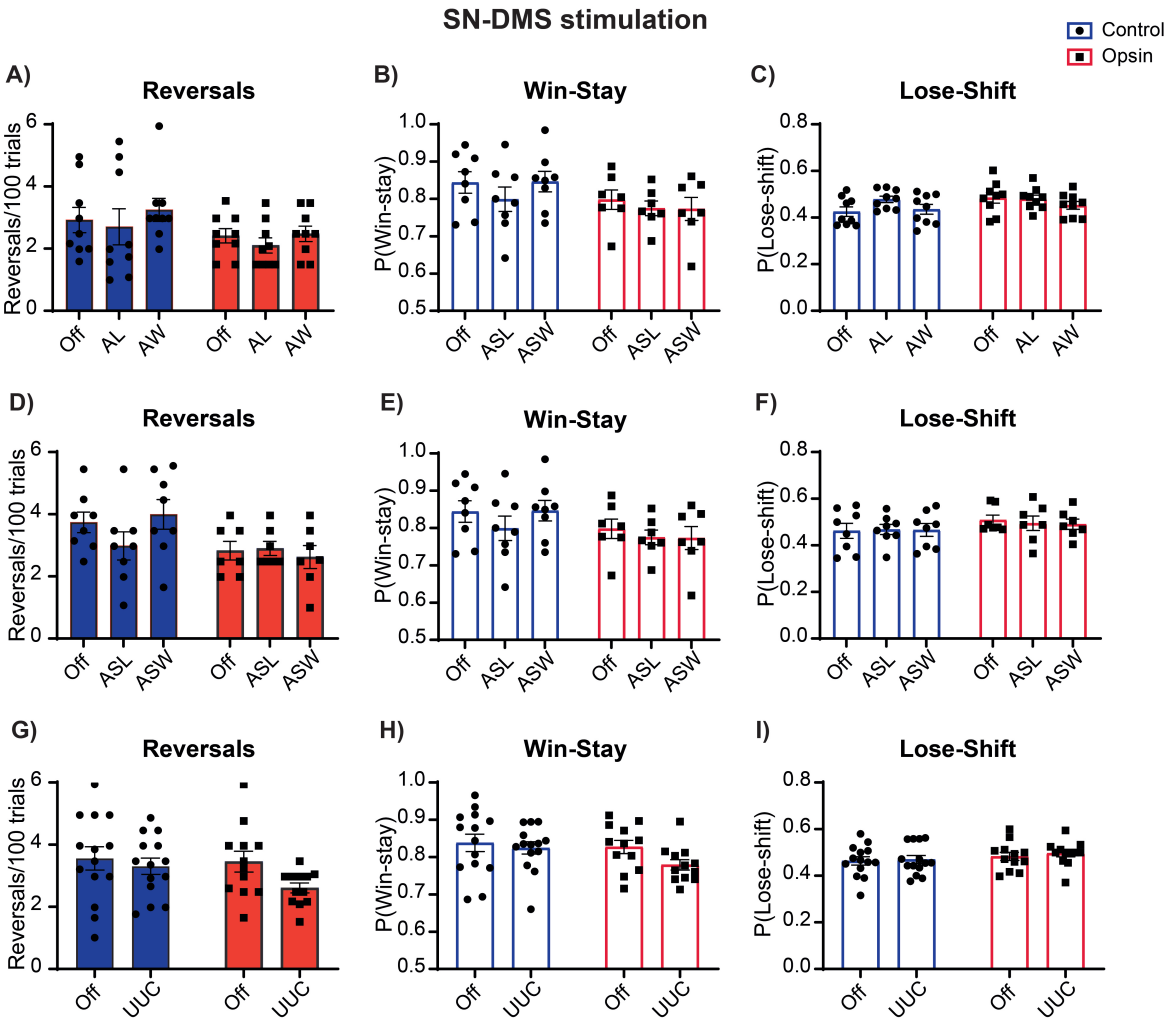


Fig. 6.4 *In-vivo* optogenetic stimulation of the nigrostriatal pathway (substantia nigra – dorsomedial striatum; SN-DMS) does not affect reversal learning. A), D), G) Reversals achieved during each session over 100 trials; B), E), H) win-stay probabilities; and C), F), I) lose-shift probabilities for the different three LSQs to test all light conditions. Light conditions: light off (Off), after lack of reward (AL), after a win (AW), after spurious lack of reward (ASL), after a spurious win (ASW) and up until making a choice (UUC). Data are shown as mean ± SEM.

For the other conditions, there was no main effect of condition (Off-AL-AW: condition: $F_{2,\,32} = 0.214$, $p = 0.310$; Off-ASL-ASW: condition: $F_{2,\,18} = 1.618$, $p = 0.227$). Nor were there significant effects of group nor a group × condition interaction in any of the tested conditions: Off-AL-AW, Off-ASL-ASW or Off-UUC (all main effect of group and group × condition interaction: $p > 0.05$).

Optogenetic stimulation of the SN-DMS pathway also did not affect latencies to respond to the stimuli, latencies to collect the reward, or omissions (Table 1).

6.4.4 Computational model parameters and simulated data

To sample latent variables influencing behaviour in the probabilistic spatial serial reversal-learning task, three different reinforced learning algorithms were used. Model 3 and model 2 provided a better fit than model 1 (average model 3: *pseudo* r^2 = 0.16, and BIC = 119.62; and model 2: *pseudo* r^2 = 0.15, BIC = 119.67, compared to model 1: *pseudo* r^2 = 0.12, and BIC = 120.71). Thus, model 3 was chosen as the best-fitting model. A fourth model was additionally tested, which included a separate learning rate (i.e. α parameter) for each lever. The aim was to investigate if learning was updated not only by the explored stimulus, but also by the non-explored option. However, this model failed to improve fit and hence was not included in the analysis.

Model 3 included four free parameters, α_{R}, α_{NoR}, β, and κ, which were fitted to the data from each session of each animal, allowing for within-subject comparisons of model performance of sessions with optogenetic stimulation against the "Off" sessions. Fig. 6.5 reports individual modelled parameters for control and ChR2-expressing rats in control light conditions and optogenetic mesoaccumbal stimulation after lack or delivery of reward. We observed a significant group $\times$ condition interaction for α_{NoR} ($F_{2,\,24}$ = 10.56, $p < 0.001$). *Post-hoc* comparisons showed this effect was marked by an impairment in α_{NoR} with light on after a win in the opsin group (p = 0.005). There was also a strong trend for a group $\times$ condition interaction for κ ($F_{2,\,24}$ = 3.360, p = 0.051). *Post-hoc* comparisons revealed the trend was due to decreased stickiness when light was on after lack of reward (i.e. AL) (p = 0.052). Whereas correlation between α_{NoR} and reversals following light on after reward (i.e. AW) was acceptable (r = 0.56), correlation between κ and reversals following light on AL was relatively strong (r = -0.65), indicating linear dependence between the computational parameters and reversal performance i.e. highlighting the correlation between the two variables.

To interrogate the validity of the winning model, the choice behaviour of agents on the PRL task was simulated based on the extracted parameters in model 3. Fig. 6.5 shows that the simulations matched the raw data for the main result: an impairment in reversals when the mesoaccumbal pathway was stimulated with optogenetics after the lack of reward, which

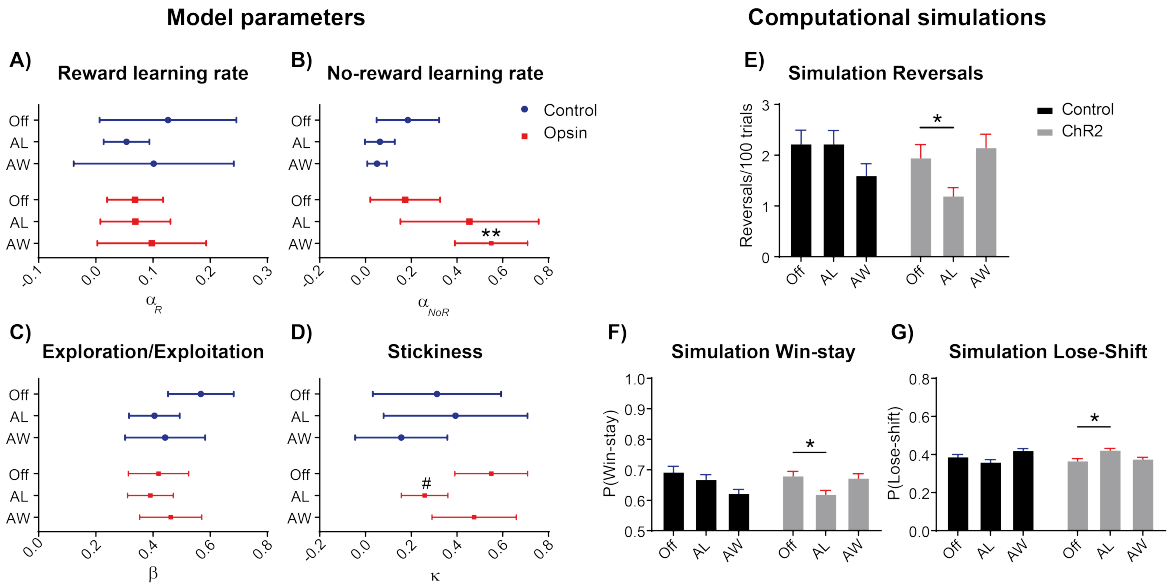


Fig. 6.5 Optogenetic stimulation of the mesoaccumbal pathway after reward omission impairs reversal learning in the PRL task. (Left panel) A) Learning rate from rewards (α_{R}) remained unaltered. B) Learning rate from lack of rewards (α_{NoR}) increased when the pathway was stimulated after a win. C) Exploration/Exploitation (β) parameter remained unaltered. D) Stickiness (κ) parameter decreased when the pathway was stimulated after a loss. (Right panel) E) Reversals, F) win-stay, and G) lose-shift probabilities analysis of data from simulated rats revealed that the behaviour of the actual rats was mostly recovered and strengthened by the winning model. Model parameters data are shown as mean ± 95% highest posterior density intervals. Simulated data (n = 40/group) are shown as mean ± SEM. ** $p < 0.01$ *vs* light off; * $p < 0.05$ *vs* light off; # $p = 0.051$ *vs* light off.

was related to a decrease in win-stay probability. The simulated data showed a significant group × condition interaction for reversals ($F_{2,\,156} = 14.950$, $p < 0.001$), win-stay ($F_{2,\,156} = 17.520$, $p < 0.001$), and lose-shift probabilities ($F_{2,\,156} = 52.570$, $p < 0.001$). *Post-hoc* analysis revealed a significant effect in the opsin group of light on AL in comparison to Off in the three variables: reversals ($p = 0.004$), win-stay ($p < 0.001$) and lose-shift probabilities ($p < 0.001$). No effects were found with optogenetic stimulation after a win.

In summary, activation of the VTA-NAcS aligned with reward omission impaired reversal learning performance. This was related to the tendency of animals to switch between levers regardless of their rewarding properties, as indicated by a decrease in the κ parameter. The κ parameter refers to the likelihood for the animals to stick to the lever to which they responded to during the previous trial. In contrast, stimulation of the SN-DMS pathway had no significant effect on any of the modelled parameters.

6.5 Discussion

In this chapter, *in-vivo* optogenetics was used to stimulate the mesoaccumbal (i.e. VTA-NAcS) or nigrostriatal (i.e. SN-DMS) pathways during a spatial PRL task. It was demonstrated that trial specific interventions neutralised the brain's response to negative RPEs when stimulation was timed with a lack of reward, or hyper-stimulated positive RPEs when timed with delivery of reward. Dissociable roles were observed for both pathways in cognitive flexibility. Whereas hyper-activation of the VTA-NAcS pathway following reward omission impaired reversal-learning performance by reducing the number of reversals achieved, no significant changes were observed after stimulation of the SN-DMS pathway. These results are consistent with the hypothesis that stimulating the mesoaccumbens pathway would impair reversal learning during reward omission.

The novelty of the present study stems from the selective stimulation of neuronal projections during specific time points in the probabilistic reversal-learning task to decipher the causal effects between neuronal activation, positive and negative feedback, and reversal-learning performance. These findings indicate that hyper-activation of the VTA-NAcS pathway aligned with reward omission impaired reversal learning, suggesting that the VTA-NAcS pathway mediates the causal link between DA neuronal signalling and reversal performance in the present spatial PRL task. This is in line with the NAcS suppressing actions to non-rewarded actions and its involvement in modulating probabilistic rather than deterministic reversal learning (Dalton et al., 2014).

Although the DMS is implicated in goal-directed behaviours and reversal learning (Castañé et al., 2010; Ragozzino, 2007; Sala-Bayo et al., 2020), stimulation of the SN-DMS had no effect on performance. This might be due to the rats in the present study being highly trained. In initial stages of reinforcement-based tasks, the DMS plays a key role to identify and apply a successful strategy. As the subjects become more familiar with the task, the involvement of the DMS diminishes, with other regions such as the DLS now serving to dominate behavioural control (Balleine and O'Doherty, 2010). Although reversals happened within each session and animals had to change their behaviour, they were familiar with this switch, thus potentially not requiring the participation of the DMS. In a stimulus-action task, whereas activation of VTA-NAcS projections induced cue approach behaviour and led to the cue becoming reinforcing on its own, activation of SNc-DMS projections induced aimless movement and did not lead to the cue becoming reinforcing (Cox and Witten, 2019). This

supports the idea that RPEs are more strongly embedded in the mesoaccumbens than the nigrostriatal pathway. Recent studies suggest that rather than being a uniform signal, DA release dynamics vary throughout different striatal subregions. Indeed, whereas phasic DA release patterns in the ventromedial striatum rapidly adapt during reversal learning (Klanker et al., 2015), DA release dynamics do not change significantly following a reversal in the dorsal striatum (Klanker et al., 2017). Instead, in this structure, small changes in DA release are observed after the onset of a visual cue, followed by pronounced DA release during lever extension or lever pressing regardless of trials being rewarded or unrewarded. This suggests that the dorsal striatum is involved in the initiation and execution of a learned operant response instead of encoding RPEs. It is also important to note that the DMS is a vast structure compared to the area that *in-vivo* optogenetics can activate, so another subregion within the DMS not affected by the light (e.g. posterior DMS) cannot be discounted from playing a role in modulating reversal learning Yin et al. (2005). Nonetheless, there was no difference in performance between posterior and anterior DMS regulating serial visual reversal-learning in rats following local infusions of D1R and D2R antagonists (Sala-Bayo et al., 2020).

After blocking the decrease in DA release that normally occurs following omission of expected reward by artificially activating the neurons targeting the NAcS, an impairment in reversal performance was found. Unexpected reward omission is considered to reduce neuronal spiking and neurotransmitter release. Previous research suggests that the observed effect could result from the combination of altered DA activity in striatal D1R and D2R. The research of Frank and colleagues (Frank et al., 2004) predicts that blocking the dip of DA into the striatum would prevent the disinhibition of the indirect D2R-expressing pathway. This would block its role controlling behaviour in a "No-Go" pathway. Hence, the hyper-dopaminergic state would hinder the suppression of actions towards the non-rewarded stimulus, and therefore impairing reversal-learning performance – as observed when stimulating VTA fibres after reward omission–. Although the lack of behavioural effects on lose-shift after hyper-activating the pathway after lack of reward runs contrary to the original reasoning, hyper-activation of VTA-NAc neurons with chemogenetics also reduced win-stay probability, with no effect on lose-shift, and decreased sensitivity to negative feedback (Verharen et al., 2018). A potential explanation could be the increased DA release acting via striatal D1R, which would activate the direct pathway controlling Go behaviour. Supporting this, administration of a D1R antagonist in the NAcS improves early reversal-learning performance, as described in Chapter 4. Whereas the improvement in

performance results from reduced activity in D1R, the observed impairment could arise from enhanced activity in D1R. In addition, striatopallidal neurons extend dense axon collaterals and target other MSNs within the same subregion in the striatum forming functional synapses (Somogyi et al., 1981; Wilson and Groves, 1980)). These connections drive lateral inhibition between the direct and the indirect pathways, so that activation of D1R- or D2R-expressing pathways could be shaping the output of the striatum by influencing on the other pathway i.e. D2R- or D1R-expressing pathways, respectively (Burke et al., 2017). As discussed in Chapters 1 and 3, D1R and D2R expression in the direct and indirect pathways in the NAcS is less segregated, hence allowing for a larger influence or competition D1R-D2R to modulate behaviour (Kupchik et al., 2015; Soares-Cunha et al., 2016). Overall, this suggests that balanced DA neurotransmission with the direct and indirect pathways is necessary for effective learning and may be regulated by a shared cellular downstream.

Our results are also in line with the observed reduced ability in rats to use loss and punishment signals to adapt behaviour following chemogenetic-induced hyper-activation of the mesoaccumbens pathway (Verharen et al., 2018). The impairment could be explained by increased baseline levels of DA, which would occlude reaching the signalling threshold that incorporates negative RPEs after an unexpected negative outcome.

Contrary to expectations, the impairment in reversal learning was related to a decrease in win-stay, not lose-shift probability. As discussed above, this could be due to the NAcS regulating reversal-learning performance via D1R following DA release. The computational model revealed no change in the learning rate AL following stimulation of VTA-NAcS neurons, but a tendency to a decreased stickiness or κ parameter, matching the impairment in reversal learning performance. Decreased stickiness in this task becomes detrimental for reversal performance, since animals would struggle to reach the eight consecutive responses to the optimal lever to achieve a reversal. Together with previous studies, our results suggest that dopaminergic projections from the VTA to the NAcS controls the likelihood of animals to repeat the previously approached stimulus while performing in a reversal-learning task.

It was also reasoned and observed that hyper-activation in the NAcS after lack of reward would impair performance. However, an improvement was expected after optogenetically activating the pathway after spurious losses, considered a real negative RPE i.e. after the lack of reward following a response to the optimal lever, which would have made animals shift less to spurious negative feedback. However, these trials only appear 20% of the times on the optimal lever, representing a minority in comparison to non-spurious-feedback trials. Their

low number throughout the task may be insignificant to modify overall performance when timed with optogenetics light. Less surprising was the lack of effect AW or ASW, since the stimulated pathway that was already being activated by the animal's experience. In this case, it could have been expected that the hyper-activation would increase the burst of neuronal spiking and facilitate reaching the threshold to act as a teaching signal in positive RPEs, or increase the rewarding properties of the reward, in both or either case leading to improved performance. Since no effect was observed, it is reasonable to assume that this is not the mechanism underlying positive RPEs processing, or that the system reached ceiling levels. To discount the latter possibility, it would be of interest to test these conditions using an inhibitory opsin.

We did not observe any effect on the "up until choice" condition either i.e. when the pathways were stimulated from the beginning of the trial until animals responded to the levers. I speculated that stimulation of mesoaccumbal cells during specific time points could strengthen reward expectancy signals, which would hinder switching to a correct response after an error. However, our results are in line with previous observations in rats experiencing local optogenetics inhibition of the NAcS: no effect during action selection, only if the light was present during reward feedback (Aquili et al., 2014).

It could be argued that lack of significant results could originate from low statistical power. However, while group sizes were not as powerful as planned, their power ranged from 0.60 to 0.93. These values stay between, or beyond, the acceptable power range of 0.60-0.80 (Cohen, 1988). Thus, the probability of presenting a type II error is rather low, especially considering that the pathway with lower power, the VTA-NAcS pathway, induced behavioural effects when being stimulated by *in-vivo* optogenetics.

The computational model also revealed a reduction in learning rate from non-rewarded trials (i.e. α_{NoR}), surprisingly when the mesoaccumbal pathway was activated AW i.e. after the delivery of reward. This could explain the trend for reduced lose-shift probability during this same condition. Notably, the simulated data shown in Fig. 6.5 provided two additional insights into our behavioural data. First, it confirmed the trend in win-stay during the AL condition (i.e. after reward omission), and it revealed a significant decline in lose-shift probability, which was in line with our original hypothesis and what the behavioural data suggested. Second, they contradicted the trend in win-stay during the AW condition, suggesting it was an artefact and hyper-activation after positive RPEs does not alter flexible performance, or alternatively the model is not accurately capturing latent processes. Further

research (e.g. replicating this experiment with a larger n), would be needed to corroborate this conclusion. Nonetheless, win-stay and lose-shift probabilities have been broadly used to measure learning from positive and negative feedback, respectively; but they are indirect measures, and both depend on the integration of positive and negative RPEs. The computational model in this study, as in previous studies (Verharen et al., 2019, 2018), shows that win-stay and lose-shift probabilities are not a straightforward measure of learning from positive and negative feedback, and are instead more strongly linked to the stickiness parameter. A more direct approach, like the VPVD task (Alsiö et al., 2019), could reveal new insights for our understanding of positive and negative feedback encoding, as shown in Chapter 3.

In the present study, a role for non-DA VTA neurons cannot be excluded. Although the virus was mainly expressed in TH^+ neurons, it was not restricted to TH^+ neurons and was therefore also present in non-TH^+, including $GAD67^+$ and $VGLUT2^+$ neurons. The observed outcomes could therefore result from the combined activity of DA, GABA and glutamate. A broad and convincing body of work has related DA to reinforcement learning, but a recent study reported that glutamate release from VTA neurons projecting to the NAc is sufficient to promote reinforcement independent of concurrent DA co-release, suggesting a DA-independent mechanism (Zell et al., 2020). Expression of VGLUT2 in VTA, but not SN neurons, is consistent with optogenetic stimulation in the ventral striatum eliciting excitatory responses, whereas activation in the dorsal striatum produces very weak or no responses (Mingote et al., 2015). In addition, DA terminals co-release glutamate preferentially in the ventromedial, not the dorsal striatum (Mingote et al., 2019; Stuber et al., 2010; Tecuapetla et al., 2010). Since the NAcS receives more glutamatergic projections than the DMS and showed larger co-expression of VGLUT2 and ChR2 positive neurons, it is plausible that the observed effects after manipulating this region were due to glutamate modulation, not DA – which is consistent with a lack of effects when targeting the DMS. However, VTA-VGLUT2 fibres predominantly innervating the medial shell contribute to aversive learning (Qi et al., 2016), which would most likely have contributed to an improved rather than impaired performance in this study, due to an increased ability to use negative feedback to adapt behaviour. In addition, intracranial self-stimulation of VTA-NAc glutamatergic terminals reinforces operant behaviour, suggesting an improvement in reversals that was not observed in our results. That said, our results are in line with the improved performance in behavioural flexibility after inhibiting NAcS DA neurons (Aquili et al., 2014) and enhanced response switching induced by systemic amphetamine (Evenden and Robbins, 1985). Therefore, further research should be conducted to establish if – or up to what point – the decrease

in reversal learning performance after activating the mesoaccumbal pathway is DA- or glutamate-dependent.

A smaller proportion of neurons co-expressed the virus and GAD67. Stimulation of VTA-GABA projections to the NAc have been shown to enhance stimulus-outcome learning, likely by inhibiting cholinergic neurons in the striatum (Brown et al., 2012). However, increased stimulus salience would result in improved performance, unlike the deficit in reversal learning observed in the present study, as it could not explain the lack of decreased reward consumption if this was delivered. Hence, combined with the finding that GABA neurons were the minority of neurons expressing the virus, it is unlikely the observed deficit in reversal learning was the result of increased GABA signalling in the NAcS.

These results open an interesting and little explored dimension to understand the mechanisms underlying reinforcement learning and cognitive flexibility. Most studies have focused on studying the role of DA encoding RPEs by using genetically modified animals, e.g. TH::Cre rodents for DA studies, but an increasing body of findings is proving this strategy insufficient, since TH-expressing neurons co-release DA with other neurotransmitters, such as glutamate or GABA. Therefore, a more sophisticated approach should be considered. For example, it would be informative to measure the different neurotransmitters released simultaneously with neuronal activation, or use conditional knockouts combined with a virus-based CRISPR/Cas9 approach to disrupt the release of certain neurotransmitters e.g. DA, while controlling for developmental adaptations (Zell et al., 2020).

Another factor to bear in mind is the potential reinforcing effect of light applied *via* optogenetics, which could be rewarding in itself. However, in similar optogenetics conditions, Steinberg and colleagues questioned the same limitation and assessed the rewarding properties of light-induced neuronal activation (Steinberg et al., 2013). They proposed that if the light itself were rewarding, then animals would need to learn two parallel associations: optogenetics light and natural reward (i.e. sucrose pellet in this thesis). If this was the case, two separate independent prediction errors should be computed, which could explain the phenomenon of blocking (i.e. impaired association between CS and US due to the presence of a second CS). They found that that their two cues interacted and competed for associative strength when being blocked, hence it was unlikely that the light was forming a parallel association. In addition, when paired with reward (e.g. in rewarded trials “after win”), the light could increase the salient value of the natural reward, or of the stimulus associated with such reward, leading to increased preference or dissociation from the non-rewarded stimulus.

However, performance in reversal learning did not improve when either the mesolimbic or nigrostriatal pathways were stimulated timed with reward delivery; neither Steinberg and colleagues found increased value or preference for paired rewards (Steinberg et al., 2013). This suggests that optical stimulation was not sufficient to compete with or enhance the properties of the natural reward.

Overall, the present research expands and supports the role of midbrain-striatal circuit encoding RPEs as teaching signals to allow learning. For the first time, it demonstrates the link between hyperactivity within the VTA-NAcS pathway and performance in reversal learning when negative outcomes are encoded. Our work adds to a scarce but increasing body of literature establishing and characterising the causal link between DA signalling during RPEs and reward learning (Aquili, 2014; Chang et al., 2015; Steinberg et al., 2013), and highlights a novel mechanism (i.e. increased response switching) for impaired reversal learning after feedback-specific aberrant DA signals. Further understanding the processing of learning from negative feedback is relevant for patients with major depressive disorder or OCD, who show an accentuated bias towards negative feedback (Clark et al., 2009; Elliott et al., 1997; Hales et al., 2014), and in the latter reduced response perseverance (i.e. stickiness) (Kanen et al., 2019).

6.5.1 Conclusions

The present study investigated the causal link between positive and negative feedback, DA signalling in mesoaccumbal projections, and cognitive flexibility. Using a spatial probabilistic reversal-learning task and *in-vivo* optogenetic stimulation, dissociable roles of the mesolimbic and nigrostriatal pathways in reversal learning were demonstrated.

We provided evidence that only the VTA-NAcS pathway computes negative feedback with stimulation reducing repetition of the just-performed response (i.e. reduced 'stickiness'). In contrast, no significant changes were detected following stimulation of the SN-DMS pathway. These findings demonstrate with an unprecedented timeframe, a causal link between activity in the VTA-NAcS pathway and reversal learning performance when negative outcomes are processed. They add to a scarce but increasing body of literature establishing and characterising the causal link between DA signalling during RPEs and reward learning (Aquili, 2014; Chang et al., 2015; Steinberg et al., 2013), and highlight a novel mechanism

for impaired reversal learning after feedback-specific aberrant DA signals (i.e. increased response switching).

This work enhances our understanding of the psychological and neural mechanisms of cognitive inflexibility and may have relevance for the aetiology and treatment of brain disorders associated with impaired cognitive flexibility.

Chapter 7

General Discussion

7.1 Summary

This thesis spans a range of investigations into the neural substrates of behavioural flexibility and their relevance in shaping learning from positive and negative feedback. Behavioural flexibility refers to the ability to adapt behaviour to changes in the environment. Although it is a crucial ability for wellbeing and success in daily life, it is impaired in a wide range of neurological and neuropsychiatric disorders. Optimal flexibility depends in part on DA neurotransmission in the cortico-striatal loop circuits, but the precise mechanism and brain loci underlying the modulatory effects of DA on behavioural flexibility are unclear.

In this thesis, cognitive flexibility was inferred in experimental rats by evaluating their behavioural performance on a set of both spatial and visual reversal-learning tasks involving a simple discrimination between rewarded and non-rewarded stimuli. During reversal, subjects must adapt and respond to the formerly non-rewarded stimulus whilst ignoring the initially rewarded stimulus. Successful performance on these tasks requires behavioural shifts, which encompass distinct psychological processes, including: (1) responding to positive and negative feedback, which shape the learning process, (2) flexibly disengaging from previous strategies by inhibiting previous behaviour, and being goal-directed to find a new strategy and gain rewards. This is then followed by forming habits that allow for fast and automatic responses lessening the cognitive demand, and (3) creating new reward-related associations, via Pavlovian or instrumental conditioning.

The research described in this thesis aimed to make headway in providing refined information of the neural, psychological and computational processes underlying cognitive flexibility to understand how flexible decision-making is modulated and to inform the aetiology of neuropsychiatric and neurological disorders that are associated with cognitive inflexibility.

7.2 The neural substrates of reversal learning: hypothesis testing

The overarching hypothesis of this thesis was that DA modulates reversal learning performance by signalling positive and negative RPEs, presumably within the direct (rewarded) and indirect (non-rewarded) striatal output pathways, respectively, and that these error signals are dissociably regulated by distinct striatal subregions and DA receptor subtypes.

I started by examining which stages of reversal learning learning (i.e. early, mid or late (Jocham et al., 2009)) are regulated by striatal dopaminergic receptors. In previous studies, presumed increases in DA transmission in the ventral striatum in pathological conditions have been associated with disrupted reversal learning (Cools et al., 2007; Dagher and Robbins, 2009). It was hypothesised that reduced DA activity in the NAc would improve performance in reversal learning tasks, and investigated which DA receptor would mediate such effect: D1R or D2R. In **Chapter 3**, a D1R antagonist, SCH23390, and a D2R antagonist, raclopride, were infused into the NAcS while animals performed a serial visual reversal-learning task. As predicted, D1R and D2R antagonism improved reversal learning. This was selective for D1R during the early phase, when animals tended to persevere in the formerly rewarded stimulus, which is now non-rewarded. Such receptor and phase specificity refines previous reports demonstrating that changes in NAc DA signalling alter behavioural flexibility (Haluk and Floresco, 2009; Verharen et al., 2019, 2018), and that different striatal subregions play a complementary role in modulating reversal learning (Sala-Bayo et al., 2020). Such mechanisms could be influenced by learning form positive or negative feedback, which are differently presented throughout the task i.e. during initial reversal animals would perseverate on the previous CS+, now CS-, hence receiving more negative feedback, which would switch to positive feedback as animals learn the task and respond to the current CS+. In addition, the present research was performed using touchscreen paradigms, which hold high translational value (Oomen et al., 2013).

Following Frank's model of the basal ganglia, I hypothesised that hyperactivation of D2R would block the signalling induced by dips of DA neuronal activity that are produced after negative RPEs i.e. after receiving less reward than expected. This hypothesis was tested by systemically administering the D2R agonist, quinpirole, while rats performed the visual VPVD task (**Chapter 4**), or the spatial PRL task (**Chapter 5**). In both cases, an impairment in reversal learning performance was found in terms of a decrease in the percentage of correct choices (**Chapter 4**) or in the number of reversals achieved (**Chapter 5**). As expected, in **Chapter 4**, this deficit was caused by blunted sensitivity to negative feedback. This extends previous work where D2R agonism impaired reversal learning across species, including rats (Boulougouris et al., 2009), monkeys (Smith et al., 1999) and humans (Mehta et al., 2001), and with the modulation of reinforcement learning by regulation of learning from negative events (Cox et al., 2015; Frank et al., 2004). Whereas this first hypothesis was based on the role of D2R in striatopallidal neurons, behaviour could also be modulated by pre-synaptically located D2R. In **Chapters 4** and **5**, I used a wide range of doses of the D2/3R agonist quinpirole: low doses were expected to act primarily on pre-synaptic receptors, whereas higher doses were expected to also activate post-synaptic D2R due to their different effect on locomotion in rats (Eilam and Szechtman, 1989). In **Chapter 5**, I also tested the effects of A2AR antagonist KW-6004, as a probe for post-synaptic behavioural effects. I found that quinpirole-induced impairment was dose-dependent, and KW-6004 replicated the deficit in performance, which, in addition, was counteracted by the D2R antagonist, raclopride. These results suggest that the effect of systemic quinpirole in reversal learning was mediated by post-synaptic D2R, which is consistent with our hypothesis and with previous research examining the role of the striatal indirect pathway in feedback modulation (Frank et al., 2004).

I then examined which brain regions were involved in regulating the observed effects in behaviour. Phasic DA release into the ventromedial striatum has been reported to predict individual differences during reversal learning performance in a manner that resembles the changes in DA signal used to correct inaccurate reward predictions (i.e. RPEs) (Klanker et al., 2015; Schultz et al., 1997). Thus, it was hypothesised that D2R in the NAc would play a key role in modulating reversal learning, which could account for the results observed with systemic manipulations. In addition, the two main subregions in the NAc, i.e. the NAcC and NAcS, which are broadly described to display complementary roles in behaviour (Floresco et al., 2006), were predicted to mediate dissociable contributions to reversal learning. Hence, in **Chapter 4**, I investigated the effect of D2R agonism in the NAcC and

NAcS. Additionally, I used an A2AR antagonist, ZM-241385, to probe if behaviour was modulated by pre- or post-synaptic D2R. The highest dose of quinpirole into the NAcC (10 μg/side) induced the impairment predicted by our systemic study. While I expected to obtain a similar result with lower doses, I instead found an improvement in reversal learning performance. This enhancement was related to increased learning from positive feedback when quinpirole was infused into the NAcS and from negative feedback when the same compound was infused into the NAcC (at doses 0.3 and 3 μg/side in both regions). In addition, whereas the A2AR antagonist replicated the effects of quinpirole in the NAcC, indicating that the change in behaviour was modulated by post-synaptic D2R, the A2AR antagonist had no significant effects when infused into the NAcS. This suggests that the dopaminergic modulation of reversal learning within striatal subterritories is much more complex than Frank's model of the basal ganglia would suggest. Although increases of DA in the NAc following L-DOPA administration in Parkinson's patients disrupt reversal learning (Cools et al., 2007), low levels of DA D2R binding in OCD patients (Denys et al., 2004), and low DA responsiveness in substance use disorder patients (Ersche et al., 2011), also predict impairments in reversal learning. This suggests that increased DA signalling in both cases would improve performance, as indicated by our local, but not systemic, NAc infusion results in **Chapter 4**.

I next interrogated the hypothesised causal link between dips of DA induced during negative RPEs and changes in reversal learning performance. I hypothesised that hyper-activation of mesostriatal neurons timed to coincide with reward omission (i.e. negative RPE) would interfere with learning by preventing the natural dip that acts as a teaching signal and would therefore impair reversal learning. To test this hypothesis, in **Chapter 6**, I used *in-vivo* optogenetics to examine whether reversal behaviour would be affected by activation of the VTA-NAcS or the SNc-DMS – counteracting the signalling dip from a negative RPE – either upon presentation (or absence) of a reward, or in the period just before making a choice.

While increased signalling in the pathway from the VTA to the NAcS timed with omission of reward impaired performance in the spatial PRL task, no effect was observed in other conditions or when targeting the pathway from the SN to the DMS. This is consistent with our hypothesis and previous literature reporting DA changes in the NAc following reversal (Klanker et al., 2015), as well as the encoding of negative RPEs within the striatum *via* the striatopallidal pathway (Cox et al., 2015; Frank et al., 2004). While I expected to observe blunted learning from negative feedback caused by a decrease in lose-shift probability, win-stay was reduced. Surprisingly, this same effect was found when systemic

quinpirole was injected during the PRL task, whilst the VPVD task clearly showed the impairment was related to blunted learning from negative feedback, not positive feedback. Computational modelling in **Chapter 6** showed that win-stay and lose-shift probabilities do not necessarily correspond with learning from rewards or lack thereof, respectively, indicating that a reframing of the meaning of win-stay and lose-shift probabilities is necessary, as well as more accurate tasks for assessing feedback sensitivity. The VPVD task perhaps makes some headway in assessing such sensitivity.

Overall, several hypotheses relating to impaired cognitive flexibility and the underlying neural underpinnings have been tested, with some providing unexpected findings. While it was found that hyperdopaminergic states in the ventral striatum affect learning from positive and negative feedback, Frank's model of the basal ganglia could not explain the behavioural results, suggesting that this model better reflects the systemic mechanisms of reversal learning modulation, or that its locus resides in another brain region e.g. the dorsal striatum, or both.

These studies span our understanding of the neural mechanisms underlying cognitive inflexibility and refine our understanding of the brain regions and receptors recruited at different stages of learning.

7.3 Cortico-striatal circuits in reinforcement learning: from Pavlovian to instrumental

Dissociating which brain subregions and receptors are involved and how they interact to regulate behaviour is critical to understand reversal learning and its underlying mechanism. Reversal learning is modulated by cortico-striatal circuits.

Previous work in our laboratory also showed that antagonising D2R in the DMS impaired the mid phase of reversal learning, which is proposed to be the period when animals have disengaged from the previous strategy and begin to adopt a new goal directed behaviour (Sala-Bayo et al., 2020). The DMS (equivalent to the caudate nucleus in primates) belongs to the so-called associative loop, which is connected to various associative cortical areas, parietal cortex and the PL from the PFC. This loop is involved in orientation, attention, affordance processing, response inhibition, and working memory, among other putative functions (Hikosaka and Watanabe, 2000; Mannella et al., 2013). Due perhaps to a combination of

these functions, the DMS is involved in learning and storing A-O associations and goal-directed behaviours (Yin and Knowlton, 2006; Yin et al., 2005), as our results suggested. Following these observations, in **Chapter 6** I investigated if the effect was causally linked to SNc-encoded RPEs while rats performed in a spatial PRL task. However, I did not find a change in behaviour when hyper-activating nigrostriatal neurons targeting the DMS, in any of the conditions investigated. This suggests that, although the DMS is involved in cognitive flexibility by mediating goal-directed behaviours and A-O associations, the role of nigrostriatal neurons in reinforcement of instrumental responses is limited or can be compensated by other pathways. Indeed, Keiflin and colleagues found that neurons from the SNc to the DMS are important for instrumental conditioning but their involvement is less than mesoaccumbal projections, which also participate in outcome predictive learning (Keiflin et al., 2019).

Of particular interest for this thesis is the *limbic loop*, whose role in reversal learning is less clear. In rats, it is formed by the NAc in the ventral striatum, which is connected to the agranular insular cortex, PL and IL (especially the NAcS). In **Chapter 3**, I observed that D1R antagonism into the NAcS selectively improved reversal learning in the early phase. The limbic loop integrates information about reward, context and motivational drive to guide behaviour (Mogenson et al., 1980), and develop goal-directed behaviours. This is especially important during early stages of learning (Dalley et al., 2004; Ragozzino, 2007). In this context, the NAcS is involved in extinction of reversal when the associated behaviour is gradually suppressed (Peters et al., 2008). This process seems to be controlled by small dips in DA signalling during negative RPEs caused by responding to the former CS+, now CS-. Therefore, antagonising D2R as in **Chapter 3** could reduce DA neurotransmission in striatopallidal neurons. NAcS activity would then better detect negative outcomes and reduce the response to the previous rewarded stimulus while enabling a search for responses receiving positive-feedback, ultimately improving performance.

I also observed that the NAcS-led improvement was specific to D1R, not D2R. The role of D1R in the NAcS might also rely on the function of striatal loops, which transfer information and contribute to "Pavlovian-Instrumental-Transfer" (i.e. PIT). Striatal projections are the main input into midbrain DA cells (Lanciego et al., 2012). The descending striatonigral pathway is organised in an inverse ventral/dorsal topography. Once in the midbrain, information is processed and these cells send nigrostriatal or mesolimbic projections exhibiting an inverse dorsoventral and mediolateral arrangement. Importantly, this creates a feedforward organisation based on spiral inputs and outputs between the striatum and the midbrain so

that information can be carried from one striatal region to the other – in particular from the ventral to the dorsal as NAcS, NAcC, DMS, DLS -, and back to the cortex *via* the thalamus (Haber, 2016).

In PIT, a conditioned stimulus that has been previously associated with reward through Pavlovian conditioning alters motivational salience and influences instrumental responding, a transition that is enhanced by training. The NAcS triggers Pavlovian conditioning and hence PIT. The NAcS acts as the gateway to pass reward properties associations to the DMS, and to the DLS, then contributing to the formation of goal-directed actions and habits (Gruber and McDonald, 2012; Mannella et al., 2013). Thus, modulating activity of the NAcS – probably the one of the initial regions in the striatum to process rewarding salience of the stimuli and their associated outcomes –, could accelerate processing in early stage of reversal learning as observed in **Chapter 3**, and reach dorsal structures more efficiently in subsequent stages. In previous studies, I observed that the DMS is involved in the mid phase of reversal learning, coming directly after the initial early phase (Sala-Bayo et al., 2020), and the DLS is implicated throughout the task, including the late phase (Sala-Bayo et al., 2020), while the NAcS and NAcC participated in early stages (**Chapter 3**; (Sala-Bayo et al., 2020)). In agreement with the concept of the transfer of information from the NAc to the DMS, in **Chapter 4**, I found that infusions of D2R agonist improved performance in the VPVD reversal-learning task from session 4 out of 10, i.e. around the mid phase of the task. This effect was selective for learning from positive feedback. Similarly, chemogenetics inhibition of the mPFC-NAcS pathway improved flexibility during early reversal learning by reducing perseverative responding (Milton et al., 2020). Thus, enhanced Pavlovian conditioning processed in the NAc, and its transfer to the DMS could explain the observed improvement in learning once animals disengaged from the previous strategy and are developing a new approach by being goal directed (i.e. session 4 in the VPVD task), engaging the DMS, and better approaching the CS+ (and avoiding the CS-). This is supported by the plateau in performance reached at later stages, suggesting that animals' level of reversal learning is not necessarily better reaching higher levels of performance (i.e. closer to 100%), but that they can solve the task faster.

While there are overlapping functions between the NAcC and the NAcS, dissociations of their role also exist (Dalley and Robbins, 2017). The NAcC generates appropriate approach behaviours towards task-related objects in a flexible manner (Day et al., 2006; Nicola et al., 2004a; Parkinson et al., 2000), is involved in assigning behavioural salience to stimuli (Berridge and Robinson, 1998) and mediates effortful responding (Salamone et al.,

2009). In **Chapter 4**, I observed that relatively low doses of the D2R agonist quinpirole improved reversal-learning performance by enhancing learning from negative feedback. The A2AR antagonist ZM-241385 produced the same improvement, indicating that our findings may have been dependent on post-synaptic D2R-expressing MSNs. However, this seems counterintuitive considering the role of the NAcC in behaviour and the expected modulation of D2R of reversal learning according to the basal ganglia model (Cox et al., 2015; Frank et al., 2004). This effect might highlight that the role of these regions does not only rely on their direct modulation of reversal learning, but on influencing other regions such as the dorsal striatum or neurons such as striatonigral MSNs *via* parallel and integrated circuits to control and adapt behaviour. Together with lateral inhibition (Salery et al., 2020) and co-expression of receptors in striatal MSNs (Aizman et al., 2000), discussed in **Chapters 3** and **4**, spiral mesostriatal loops highlight the integration of cognition and reinforced learning throughout multiple brain regions and cross-talk interactions. Thus, our results could stem from a competition with different striatal neurons over the control of animals' cognitive and motor resources (Lauwereyns et al., 2002; Nicola et al., 2004b), and the observed improvement could be modulated by D2R-expressing MSNs acting on D1R-expressing neurons.

Conversely, in **Chapter 4**, the highest dose of quinpirole infused into the NAcC induced an impairment in reversal learning by blocking learning from negative feedback. A decrease in levels of performance was also observed with systemic quinpirole in **Chapters 4** and **5** (although in **Chapter 5** learning from negative feedback was not assessed). Similar to our results, Verharen and colleagues found an impairment in reversal learning when a high dose of quinpirole (5 µg/side) was infused into the ventral striatum while rats performed a spatial PRL task (Verharen et al., 2019). Effects of quinpirole in this thesis were therefore dose-dependent and suggest that when the system is saturated in the NAcC, direct and indirect pathways regulate behaviour as proposed by Frank's model of the basal ganglia (Frank et al., 2004). This emphasises the plausibility of strong DA signals in the ventral striatum being transmitted to the dorsal striatum, where regulation of reinforcement learning is better constrained by D1R and D2R modulating learning from positive and negative feedback, respectively. An example of transferring information from the ventral to the dorsal striatum is observed following infusions of amphetamine into the NAcC (Wyvell and Berridge, 2000). Increasing local DA levels with amphetamine potentiates PIT for food rewards, which, as discussed, is a process based on transferring information from the NAc to the DMS and DLS (Wyvell and Berridge, 2000).

These findings demonstrate the role of different regions in the striatum having dissociable roles and sensitivities to different receptors, which suggest a competing or complementary role in modulating reversal learning.

7.4 Learning from positive and negative feedback and clinical implications

Reversal learning tests require the subjects to discriminate between different options and associate each option with specific rewarding properties, which shift after achieving discrimination. Solving the task (i.e. disengaging from the previous rewarded stimulus and responding to the originally unrewarded stimulus) is often studied as a single-faceted cognitive process. However, subjects can develop new strategies by learning from positive or negative feedback. That is, subjects can solve the task by either approaching the rewarded stimulus or avoiding the negative. A key factor to expand our understanding of the neural mechanism underlying cognitive flexibility is therefore to comprehend how brain regions integrate and process sensitivity to positive and negative feedback.

In **Chapter 4**, I found that enhanced overall activity of D2R led to decreased sensitivity to negative feedback in the VPVD task. Intra-NAcS D2R agonism improved reversal learning performance by promoting learning from positive feedback. Similarly, in **Chapter 6** I found that manipulating input signalling from the VTA to the NAcS decreased win-stay probability. Win-stay probability has traditionally been used to assess sensitivity to positive feedback, as lose-shift is accepted to reflect sensitivity to negative feedback (Rygula et al., 2018), although other studies have suggested it is not a reliable measure of feedback sensitivity (Alsiö et al., 2019; Verharen et al., 2019, 2018). Ventral striatal regions have been especially involved in learning the value of positive feedback linked to a particular stimulus (Rygula et al., 2018). In rats, Dalton and colleagues observed that manipulating the NAcS reduced win-stay in the PRL task (Dalton et al., 2014). However, their effect was induced by inactivating the NAcS, which also impaired overall performance, as our pharmacological study in **Chapter 4** would have predicted. In a subsequent study, inactivation of the mOFC, which projects to the medial NAc (i.e. NAcS), reduced sensitivity to positive and negative feedback during the initial phases of discrimination (Dalton et al., 2016).

Conversely, when the D2R agonist was infused into the NAcC, it initially improved performance by enhancing learning from negative feedback, while it switched to impairing reversal learning by blocking sensitivity to negative feedback at the highest doses. The direction of the effect in the NAcC was dose-dependent. The effect induced by the highest dose matched with the impairment I found in **Chapter 4** when the D2R agonist was administered systemically, modulating more brain regions involved in reversal learning. As discussed in previous sections, this transition might reflect the integrative communication between brain regions and neuronal types modulating behavioural flexibility, including the dorsal striatum. In this line, inactivation of the lOFC (which projects to the DMS) impaired reversal stages in the PRL task by making rats shift less when responding to the incorrect stimulus, suggesting that the lOFC and its efferent projections may be particularly important in signalling negative feedback and the violations of reward expectancies (Dalton et al., 2016).

Abnormal sensitivity to feedback has been observed in patients with depression (Elliott et al., 1997; Gradin et al., 2011; Rygula et al., 2018). Depression has been broadly linked to a cognitive bias towards negative feedback. A growing number of studies suggest that, apart from hypersensitivity to negative feedback, depressive patients also show hyposensitivity to positive feedback (McFarland and Klein, 2009; Robinson et al., 2012). Several attentional and reversal learning tasks can be applied in humans to assess underlying mechanisms of cognitive flexibility and reinforcement learning (Izquierdo, 2017; Rygula et al., 2018). The PRL paradigm is also used in humans to investigate the neurochemical and neuroanatomical correlations of feedback processing. Patients with depression show cognitive inflexibility (Remijnse et al., 2013) and their negative bias is linked to aberrations within the reward system (Keedwell et al., 2005; Tremblay et al., 2002). In **Chapters 3**, **4** and **6** I found that the NAc in the ventral striatum played a critical role in modulating reversal learning and the sensitivity to feedback. Indeed, ventral striatal activity in patients with major depressive is altered, showing reduced activity in response to positive stimuli in comparison to healthy volunteers (Epstein et al., 2006; Surguladze et al., 2005).

Parkinson's disease patients tend to learn faster from negative feedback than positive feedback (Frank et al., 2004). Critically, this abnormality depends on whether or not the patient is under treatment with L-DOPA. By increasing DA levels in the brain, especially the SN, L-DOPA shifts the learning bias towards positive feedback. This disparity in performance between medicated and non-medicated Parkinson's patients has resulted in theoretical explanations that attribute these changes in the bias to adaptions in the striatal dopamine system. As seen in **Chapter 4**, increases in D2R activity in the NAc influence

positive and negative feedback, and in a dose-dependent manner can switch from inducing an improvement to an impairment in feedback sensitivity.

Similarly, OCD patients show attentional bias towards threats, although there are contradictory reports in this area. It has been observed that individuals with OCD take longer to respond to words with negative valence, and OCD washers (i.e. patients that perform excessive and repetitive washing) show greater conflict for words representing poor cleanliness, their most negative trigger (Lavy et al., 1994). OCD patients also show reduced binding to DA receptors in striatal regions (Denys et al., 2004). Although this is in contradiction with our D2R antagonist results in the NAcS from **Chapter 3**, it highlights that D2R agonism, as found in **Chapter 4**, or increased DA activity in the striatum could hypothetically improve flexible performance in subjects with low DA activity baseline, which has been observed in drug abusers (Ersche et al., 2011; Kanen et al., 2019).

Our experiments provide region, receptor and phase specificity to the modulation of behavioural flexibility and learning from positive and negative feedback. They also emphasise the importance of personalised medicine specific to the individual, as the same manipulation might have opposite effects depending on the targeted area, and while patients cannot be locally treated yet, treatment efficacy could be predicted depending on patients' neuronal impairment. Our results provide refined information on the neural circuit underlying cognitive disorders, which contributes to the understanding of how different symptoms (i.e. sensitivity bias) relate to distinct brain mechanisms.

7.5 Causal link between phasic DA and behavioural performance

RPEs are a crucial parameter in associative learning models. The sign and magnitude in midbrain DA neurons vary according to the degree to which the reward is expected (Cohen et al., 2012; Hollerman and Schultz, 1998), which manifest bilaterally, in both VTA and SN. Despite the prevalence and impact of the theory that RPE signalling by DA neurons induces associative learning, only a few studies have supported a causal role between DA-mediated RPE activity and learning about natural reward e.g. food. Indeed, this question had not been explicitly tested in the context of reversal learning. In **Chapter 6**, I aimed to address this by testing the causal link between overdoses of DA in the nigrostriatal or

mesoaccumbal pathways during feedback processing and performance in the spatial PRL task in rats. The impairment I found on reversal learning performance following hyper-activation of mesoaccumbal, not nigrostriatal, neurons at the time of reward omission is the first direct evidence of RPEs being causally related to reversal learning. This expands on the correlations found between increases in DA neuron firing, especially within the VTA (Fields et al., 2007; Swanson et al., 1997). In this line, microdialysis studies have shown that DA release is strongly correlated with behavioural activity, particularly when the release occurs into the NAcS (Freed and Yamamoto, 1985; Hamid et al., 2016). While this effect was predicted by the model described by Frank and colleagues (Frank et al., 2004), which suggests that hyper-physiological levels of DA during integration of negative outcomes blunt the natural dip in release and disrupt learning, I did not observe a change in lose-shift probabilities, canonically linked to sensitivity to negative feedback (Rygula et al., 2015). Instead, I found decreased win-stay probability, commonly related to sensitivity to positive feedback. This indicates that NAcS mediates learning from negative feedback by adjusting the approach behaviour towards the newly rewarded stimulus, originally unrewarded. However, the learning rate obtained from the computational model was not affected by the quinpirole treatment, but instead animals showed decreased stickiness, which indicates that instead of affecting how animals learned from reward or omissions of reward, optogenetic stimulation made animals less likely to remain in the same stimulus from the previous trial in general. The NAcS and its afferents from the VTA might be key in producing and modulating the teaching signal in cognitive flexibility, although RPEs might not be directly modulating learning in the context of reversal learning. This raises the question of what is the exact role of DA signalling RPEs within these regions.

First, fluctuations of DA levels are widely accepted to have two separable modes: large amplitude subsecond phasic release, and steady-state 'background' tonic release arising over a timescale of minutes (Grace, 1991). Previous literature proposes that phasic transmission may support reward learning information (including RPEs) (Klanker et al., 2015; Wanat et al., 2009), whereas tonic changes may instead modulate overall motivation or arousal (Schultz et al., 1997) or, more specifically, enable animals to exploit the reward-related information that they have learnt (Beeler et al., 2010). Since drugs alter DA levels for prolonged periods, whilst *in-vivo* optogenetics alter neuronal activity on the millisecond timescale, it could be interpreted that pharmacological results represent changes in tonic release, while optogenetics reflects changes in phasic release. This would be supported by the drug-modulated behaviour I found in **Chapter 4**, which was related to increased percentages

of optimal choices towards the positive stimulus when targeting the NAc or, in other words, exploiting positive feedback. However, the computational model did not show any difference in the β parameter or exploitation/exploration rate and I also observed changes in win-stay probability following optogenetics stimulation. Nonetheless, I did not assess tonic or phasic changes in DA in our studies, and direct evidence of covariance between tonic transmission and motivational states is lacking. Others have also observed modulation of phasic firing by changes in motivations, not only RPEs, in reinforcement learning (Phillips et al., 2003; Wassum et al., 2012), suggesting the role of phasic and tonic DA signalling is not clear-cut.

Second, the current work was strongly based on and is supportive of the concept that midbrain DA cells signal RPEs that act as teaching signals in reinforcement learning (although **Chapter 6**. DA transmission has a striking similarity to a reinforcement error signal representing the update between predicted and delivered reinforcer, and its timing is characteristic and precise ($\sim$100 ms of latency and $\sim$100 ms of duration), which suggests that the role of DA in timing with the reinforcement is crucial (Redgrave and Gurney, 2006). Although a large and compelling body of work supports the role of phasic DA as an error teaching signal, recent studies have questioned this view (Hamid et al., 2016; Howe and Dombeck, 2016; Wassum et al., 2012). DA signal persists after prolonged training and steadily increases (i.e. ramps) over the duration of each trial in certain tasks (Hamid et al., 2016; Howe et al., 2013; Kim et al., 2019). In light of these observations, it has been suggested that DA-RPE theory by Schultz (Schultz et al., 1997) could represent a fraction of a larger set of functions for DA that act as teaching signals. In other words, Schultz's RPE theory for DA may only apply to specific events.

Nevertheless, although compelling evidence indicates that DA neurons provide teaching signals, the intricacies of such signalling are unclear and little proof exists to support a causal link between both phenomena, especially in reversal learning. In this thesis, I used behavioural procedures in which learning is considered to be driven by RPEs and targeted neuronal pathways involved in reinforcement learning with high temporal precision. Hence, the data presented in **Chapter 6** are the first approach to investigate the causal relationship between phasic DA signals timed with reward omission during reversal learning and reversal performance, and emphasise the dissociable role between nigrostriatal and mesoaccumbal projections.

7.6 New therapeutic approaches

Cognitive inflexibility in neuropsychiatric and neurological disorders is often treatment-resistant or even worsens with pharmacotherapy. To overcome this limitation, treatment tools are expanding to non-dopaminergic strategies. Adenosine receptors antagonists are being tested in clinical trials for disorders such as Parkinson's disease, OCD and depression (Asaoka et al., 2019; Hung and Schwarzschild, 2014; Leibenluft et al., 1993). For example, caffeine, an A1AR and A2AR antagonist, has been suggested as alternative preventive and complementary treatment in Parkinson's patients, where it has been reported to improve both cognitive and motor symptoms (Prediger, 2010).

A2AR co-localise with D2R on striatopallidal MSNs and converge onto the same signal transduction downstream pathways in an antagonistic way. Selective A2AR antagonists have shown positive results as treatment for pathologies related to dopaminergic dysfunction (Hung and Schwarzschild, 2014; López-Cruz et al., 2018). In **Chapter 5**, antagonism of A2AR not only counteracted the raclopride-induced decline in performance but also left latencies or trials completed per session intact. Similarly, in **Chapter 4**, local infusion of the A2AR antagonist KW-231485 into the NAcC and NAcS improved reversal-learning performance without increasing latencies to collect the reward, as opposed to quinpirole in these same regions, which made animals slower. These findings suggest that targeting A2AR in D2R related disorders is an advantageous approach in clinics to incorporate the benefits (e.g. recovery of cognitive flexibility) of altering DA activity without its disadvantages (e.g. impaired motor control).

In individuals with neurological and neuropsychiatric disorders not only is cognitive flexibility affected, but other symptoms are also present such as lack of motivation, decreased psychomotor speed, and fatigue (Goldsmith et al., 2016; López-Cruz et al., 2018). Adenosine receptor antagonists have proved to recover symptoms related to both dopaminergic and non-dopaminergic dysfunction. However, high doses of methylxanthines (e.g. caffeine) are often poorly tolerated (Frozi et al., 2018). I also observed in **Chapter 5** that a relatively high dose of systemic KW-6002, the A2AR antagonist, impaired reversal learning, but when administered following a raclopride-induced imbalance of DA signalling, it recovered flexible performance. This indicates and supports previous literature suggesting that adenosine receptor antagonists might be particularly effective when the dopaminergic system is compromised.

Determining when patients would benefit from a dopaminergic or an alternative strategy is key to therapeutic success. Further research is necessary to understand the predominant symptomatology, neural mechanisms, and potential drug effects on each patient. The present thesis supports the therapeutic benefits of A2AR antagonism and expands our comprehension of the circumstances whereby it might improve or impair symptoms, as well as providing brain region specificity for its effect.

7.7 Methodological considerations

7.7.1 Behavioural tasks

VPVD task

The VPVD task was originally suggested by Nilsson and colleagues (Nilsson et al., 2015) and builds upon previous research aiming to dissociate between perseveration or inhibition of responding to the no longer rewarded stimulus (A-), and learned non-reward or approaching the previously rewarded stimulus (B+) (Alsiö et al., 2015; Clarke et al., 2007; Piantadosi et al., 2018). As a task, it has its strengths and weaknesses, which are discussed below.

The main strength of this task is its ability to assess how learning from positive and negative feedback impact reversal-learning performance. This is achieved by using probe trials in which it is necessary to leverage a probabilistically reinforced "neutral" stimulus against a deterministic positive or negative stimulus. Since tracking the preference of stimuli happens across all the reversal phase, it allows us to investigate feedback sensitivity as a learning curve. This is an advantage over previous tasks, like the PRL, in which feedback sensitivity is extracted from trial-by-trial changes responding to immediate reward, or lack of it, as win-stay or lose-shift probabilities. That is, the probability of animals to stay on the same stimulus after receiving a reward or shift after a loss or lack of reward. In order to behave in a win-stay or lose-shift manner, animals need to simultaneously integrate positive and negative feedback, since deciding if it is more valuable to shift or to stay on the same stimulus depends on an integrative processing of both the received feedback and the expected value of the other stimulus. In addition, as seen in **Chapter 6** and previous literature, computational modelling has revealed that there is no correlation between win-stay/lose-shift probabilities

and learning rates from rewarded or unrewarded trials, respectively (Alsiö et al., 2019; Verharen et al., 2019). The VPVD task overcomes the limitation of these probabilities since measurement of positive and negative feedback sensitivities do not require the simultaneous assumption of the rewarding properties of the explored and non-explored options. In addition, tracking stimulus preferences throughout the length of the learning curve allows for the investigation of events characteristic of specific stages in the task e.g. perseveration towards the previously rewarded stimulus, visible in early stages of reversal learning, or formation of habits, detectable throughout the late stages.

Another benefit of the task is that the dissociation of the two type of sensitivities (i.e. positive and negative) together with learning on standard trials can capture subtle changes that would not be detected if only performance as a whole was investigated. I observed this phenomenon with the mid dose of quinpirole (0.1 mg/kg), which affected learning from negative feedback in later sessions in the reversal phase but was not sufficient to affect the overall performance.

The VPVD task is not without limitations, however. Although trials in which the $C_{50/50}$ stimulus is presented are considered probe trials to investigate how animals are learning to discriminate A- and B+, it is plausible that animals are actually learning to solve three different pairs, instead (i.e. A- *vs* B+, A- *vs* $C_{50/50}$ and B+ *vs* $C_{50/50}$. To minimise separate learning, probe trials are presented less often than standard trials (i.e. A- *vs* $C_{50/50}$ or B+ *vs* $C_{50/50}$ each appear only 1 in 8 trials). This ensures that the main learning is based on the standard deterministic trials but cannot completely remove the possibility of separate learning. In addition, reversal-learning performance might be highly influenced by other processes in addition to cognitive flexibility, such as attention, locomotor activity, and motivation.

PRL task

Despite some evidence suggests that the PRL task does not provide accurate information on the subjects' sensitivity to positive or negative feedback, it has been broadly used due to its advantages in animals and its translation to humans. In animals, the PRL paradigm allows for testing effects of different probabilities of rewards (i.e. uncertainty) while assessing the effect of experimental manipulations (e.g. drug, optogenetics, and chemogenetics) as between- and within-subject factors – with the latter applicable only for reversible treatments. As seen in **Chapters 5** and **6** (i.e. PRL) in comparison to **Chapters 3** and **4** (i.e. VPVD), rats

learn the spatial PRL task faster and can perform multiple reversals within the same session, whereas in the VPVD paradigm they need more weeks to reach the discrimination criterion and subsequently multiple sessions to complete one reversal. A potential explanation for this is that in touchscreens, animals are removed from their standard habitat and context to train or test, hence potentially raising problems related to the ethological validity of the methodology. Specifically, due to the artificial nature of the setting, subjects may rely on processes to solve the task that are not standardly recruited in their natural environments. During the PRL task, animals are also tested in operant chambers, but they rely on spatial, instead of visual, abilities, which are highly developed in rodents. In addition, this dissociation in learning span might depend on the speed in which the neural circuits form S-R associations following Pavlovian conditioning within the striatum, and its transfer to other striatal regions to develop operant behaviour (see section 7.3 within this chapter) within this chapter). Hence, although assessing sensitivity to feedback with the PRL task might be misleading, this task provides a platform to test reversal learning, with other pragmatic/convenient advantages such as reduced time-frames, as it can be rapidly trained.

7.7.2 Computational modelling

The experimental work in this thesis was performed in rats, but the ultimate goal was to provide an understanding of the human brain. While human and rodent brains share structural and functional similarities, it remains unclear whether our findings have any relevance for humans. The use of analogous paradigms of reversal learning in animals and humans contributes to making the translation possible, but different aspects of behavioural responses are still challenging to reliably relate across species.

Computational modelling of behavioural data enables us to further understand behavioural results and increase the translational value of animal studies by assessing subtle changes in the employed strategy to solve the reversal-learning paradigm. The parameters that can be derived from computational models provide insight into the mechanisms that form the basis for complex and dynamic decision-making behaviour, including learning from delivery or omission of reward, and perseverative behaviour. In **Chapter 6**, the computational model revealed that animals with impaired performance were not modifying their learning rate depending on reward or lack of it, neither that they changed their behaviour by exploring or exploiting the stimuli according to their accumulated evidence of reward. Instead, they

were less likely to repeat previous choices during the "after loss" condition. This provided insight not only into the behavioural trait – and its underlying neural mechanism –, but also into the task, since, as discussed, it revealed that the traditional way of interpreting win-stay and lose-shift probabilities as sensitivities to feedback may be incorrect.

However, computational modelling is not exempt from downsides. Choosing the wrong model or data over-fitting can lead to false conclusions. Care and preparation must be taken to implement the appropriate model for each situation depending on the context of the results. Data availability might be restraining to obtain accurate and reliable results as experimental approaches use small group sizes ($n < 100$), whilst these studies would most benefit from larger datasets. Nonetheless, despite computational modelling still being relatively nascent, it has real potential as a translational bridge to inter-relate behavioural and psychological constructs in humans and other animals.

7.8 Limitations and alternative approaches

A number of limitations should be taken into account when interpreting the present results, in addition to those mentioned in each chapter. First, the D2R agonist, quinpirole, and antagonist, raclopride used in **Chapters 3**, **4**, and **5** also have affinity for D3R, D4R and serotonergic receptors (Knight et al., 2004; Millan et al., 2002). Although serotonergic receptors and D4R are less abundant in the NAc, D3R are highly expressed in the NAcS (Schwartz et al., 1994) and it is not possible to clearly distinguish if the observed effects were due to D2R or D3R actions, which could account for the unpredicted findings in the studied subregions. For example, antagonising D3R in the NAcS has been reported to increase impulsivity in rats, hence their agonism may have the opposite effect: decrease impulsivity and improve flexible performance (Besson et al., 2010). To surmount this limitation, I could repeat the experiments using a more selective D2R agonist, such as sumanirole, or assessing the role of D3R in behaviour by combining D2R agonism with selective D3R antagonists such as nafadotride (Sautel et al., 1995).

Second, in rodents, chronic administration of quinpirole induces sensitization, which generates compulsive-like behaviour (Amato et al., 2006; de Haas et al., 2011; Szechtman et al., 1998), and increased locomotion activity (Escobar et al., 2017). Sensitization to the D2R agonist quinpirole seems to be induced by reduced DA release, especially in the NAcC

(de Haas et al., 2011; Escobar et al., 2017), increased post-synaptic D2R sensitivity (Escobar et al., 2017), increased D2R binding, and reduced glucose consumption (Culver et al., 2008). Whereas quinpirole-induced compulsive-like behaviour could have an impact in reversal learning, such behavioural traits would impair animals' performance. Nevertheless, when repeatedly administering quinpirole into the NAc, I observed an improvement in performance selective for the early perseverative phase of reversal learning. In addition, a recent study by Eagle et al. (2020) showed that chronic quinpirole only increased functional, but not dysfunctional checking behaviour in an observing response task in rats, suggesting that it might not accurately replicate OCD behaviours. To control for accumulative effects in locomotion in future experiments, I could run a locomotor activity test with freely moving rats before and after the testing period (i.e. before the first, and after the last testing day).

Third, comprehending and dissecting the role of DA transmission is intricate since D2R are expressed in both pre- and post-synaptic striatal neurons, and on striatal GABAergic and cholinergic interneurons (De Mei et al., 2009). Pre-synaptic D2R or autoreceptors mediate auto-inhibition to control DA release into the cleft in conditions of high extracellular DA levels. For example, in drugs of abuse such as cocaine that block the DA transporter, pre-synaptic D2R are the only factor left to counteract the hyperdopaminergic effect, which explains why D2R play a critical role in drug abuse disorder (Centonze et al., 2002; De Mei et al., 2009). Pre-synaptic D2R are expressed in striatal neurons not only projecting from the midbrain, but also from the cortex, and are implicated in GABA, glutamate and acetylcholine release (Bamford et al., 2004; Centonze et al., 2004; Wang et al., 2006). Therefore, we cannot exclude the possibility that altered concentrations of these neurotransmitters impacted on our results.

Autoreceptors mainly belong to the short isoform of the D2R (i.e. D2S), which is more sensitive to DA than the long isoform (i.e. D2L) and interacts with different signalling pathways, whereas post-synaptic receptors are mainly D2Ls and modulate activity of striatopallidal neurons (Lindgren et al., 2003; Usiello et al., 2000). However, it has been reported that D2S can also be expressed in post-synaptic neurons and inhibit D1R responses, while acting in synergy with post-synaptic D2L (Usiello et al., 2000). Thus, the effects induced by quinpirole and raclopride in **Chapters 3**, **4** or **5** might not be exclusively mediated by D2R, but also by D1R *via* D2S, in addition to the potential interacting mechanisms discussed in the relevant chapters such as lateral inhibition. To dissociate the modulation related to different D2R-expressing neurons and their isoforms, we could use selective or conditional

knockdowns of D2R, for example using the Cre-loxP system (Sato et al., 2004), and replicate the experimental designs presented in this thesis.

Another factor to bear in mind is that performance in reversal learning may rely on other cognitive processes in addition to or instead of cognitive flexibility, such as attention, motivation or locomotion. Although throughout testing I aimed to control for these processes (e.g. using latencies to collect the reward or to respond to the stimuli), a direct approach would be necessary to weigh their influence in observed behaviour. A possibility would be to test animals' behaviour in different tests e.g. effort-related choice for motivation (Salamone et al., 2018), 5-choice serial reaction time test for attention and response control (Carli et al., 1983), or infrared detection for spontaneous locomotor activity. A better alternative would be to assess a wider range of cognitive processes within each task, for which additional task development would be necessary. Furthermore, animals are typically removed from their habitual environment to train and test, arising potential problems related to the ethological validity of the methodology. In response to this limitation, different automated systems are being developed to assess cognitive abilities while animals remain in their home cage (Aarts et al., 2015; Galsworthy et al., 2005; Redfern et al., 2017).

Finally, all the studies presented in this thesis were performed with male rats and bilateral manipulations. It is conceivable that future studies with females will reveal sex differences in the impact of DA receptors and the neural circuitry on reversal learning, as unilateral manipulations might expose compensatory, competing effects within hemispheres or reveal intricacies in regional dynamics and inter-communication.

7.9 Future directions

One prominent question that remains to be answered is how the dynamics of DA release, neuronal firing and RPEs during flexible decision-making relate to one another. To determine the link between these phenomena, tracking DA fluctuations with high temporal and spatial resolution is key. The recently developed neurotransmitter sensor dLight1 (Patriarchi et al., 2018) would allow to monitor DA modulatory signal dynamics while animals performed in a reversal-learning task. Alternatively, *in-vivo* microdialysis or electrochemical sensors could provide an understanding of DA release, but lack spatial and/or temporal resolution and cannot target specific cells of interest (Jaquins-Gerstl and Michael, 2015; Muller et al.,

2014). To better understand the DA connection between brain regions, DA release should be compared with DA neuronal activity. For this, fibre-photometry could be used to characterise the dynamics of population-defined neural activity, and test how manipulating DA activity with optogenetics affects the dynamics of DA release and the choice process. All these manipulations could be applied while animals perform in a reversal-learning task that dissociates learning from positive or negative feedback – such as the VPVD task – to differentiate the effect on positive and negative RPEs.

There is supporting evidence that DA concentrations in the striatum do not always reflect the activity of midbrain DA neurons (Mohebi et al., 2019). Unlike positive RPEs, which induce a large increase in neuronal activity, negative RPEs induce a small pause in spiking in the midbrain. Some have suggested that this small pause translates into a larger decline in DA release in post-synaptic regions, potentially mimicking the magnitude of the increased efflux following positive RPEs (Hart et al., 2014), again reflecting the fact that neuronal activity and DA release do not necessarily correlate. It is unclear how small pauses after negative RPEs are sufficient to encode teaching signals, but when inhibiting midbrain neurons with optogenetics, brief pauses proved sufficient to serve as RPE, although the observed behavioural results were modest (Chang et al., 2015). It is plausible that, instead of magnitude, negative RPEs are encoded by extended pauses in neuronal activity. Therefore, longer pauses in neuronal spiking could induce a larger effect. Testing if the duration of pauses in midbrain DA firing serves as a teaching signal in negative RPEs could be achieved by assessing how manipulating DA neuronal activity affects DA release dynamics and behaviour. Specifically, using optogenetics would allow testing of the impact of inhibiting selective VTA DA neurons for different durations and measuring the effect on behaviour and DA release with fiber photometry (Saunders et al., 2018).

The present thesis has mainly focused on systemic or ventral striatal manipulations (apart from manipulation of nigrostriatal projections to the DMS in **Chapter 6**). However, cognitive flexibility is not only modulated by the ventral striatum and greatly relies on the interconnectivity between this and other brain regions, including the PFC, the OFC, and the dorsal striatum (Hervig et al., 2019; Izquierdo et al., 2017; Turner and Parkes, 2020). Further elucidation of cortico-striatal circuits using new methodology such as optogenetics and chemogenetics, or the recently developed behavioural tasks (Alsiö et al., 2019) could contribute to target specific projections from different subregions in the cortex to the basal ganglia. In addition, there are many understudied circuits within each region and it is likely that subregions not only receive inputs from and send outputs to other subregions, but that

they act as cooperative and interconnected hubs before communicating with other parts of the brain. Thus, microcircuits could be a potential avenue for future research.

If preclinical research aims to provide insight into the neural mechanisms underlying cognitive disorders in humans, it is of vital importance that behavioural parameters used in animal studies appropriately relate to and reflect the relevant system involved in patients. Further efforts should be placed on developing new or optimising current tasks to (1) dissociate parameters and processes that might be underlying behaviour to refine the results and our understanding; (2) provide more naturalistic approaches to animals' behaviour, which in turn contributes to the last point; (3) assess behaviour with high construct and validity, and provide translational power to transfer findings in animals to humans with as little ambiguity as possible.

Finally, computational modelling can increase translational value of preclinical studies by assessing subtle changes in behavioural strategies employed by animals and humans when performing in the task, as discussed in section 7.7 and observed in **Chapter 6**. These models can also relate behavioural findings with neurophysiological data by encoding the dynamic nature of neurons, as shown by the drift-diffusion model (Wiecki et al., 2013). Thus, future preclinical and clinical studies should ideally include computational tests as part of their systematic analysis, where appropriate. In addition, novel statistical approaches arising from machine and deep learning could be particularly informative in identifying complex patterns or relationships not captured by traditional methods, leading neuroscience to a highly promising future that is about to come.

7.10 Conclusions

The findings reported in this thesis collectively show that behavioural flexibility is differentially affected by sensitivity to positive or negative feedback, and its modulation relies on distinct neural circuits, striatal subregions, and dopamine receptors.

By combining in different experiments systemic, local pharmacology and optogenetic approaches while animals performed in three different reversal-learning tasks, this thesis revealed several novel mechanisms that underpin behavioural flexibility. First, I provided experimental evidence that D2R mediate learning from negative feedback, potentially by acting on striatopallidal neurons. Second, I showed that systemic D2R agonism impairs

reversal learning, whilst hyper-activation of D2R in the NAc leads to an improvement, suggesting that different regions act in synergy or compete to modulate behaviour. Third, NAcC and NAcS utilise dissociable mechanisms to regulate flexibility: whereas the NAcS mediates sensitivity to positive feedback potentially *via* pre-synaptic D2R, the NAcC mediates negative feedback possibly *via* post-synaptic receptors. Fourth, levels of D2R agonism in the NAc alter the balance of DA function and can switch from inducing improvements to impairments in reversal learning. Fifth, although the effects of manipulating the NAc might be expressed in later sessions in the VPVD task, the NAc modulates performance in early stages of serial reversal learning, suggesting it plays a critical role in disengaging from previous strategies to develop a new approach. Sixth, hyper-activation of the mesoaccumbal, not nigrostriatal, pathway impairs reversal learning when timed with lack of reward, not with the delivery of reward or up until making a choice. Finally, by comparing our results from the VPVD task with those obtained in the PRL task and analysing the latter with computational models, I demonstrate that the canonical measure of positive and negative feedback in animals and humans (i.e. win-stay/lose-shift probabilities in the PRL task) is not necessarily accurate and can, in fact, be misleading.

The results from this thesis demonstrate with unprecedented detail how striatal DA modulates feedback sensitivity, bringing the field closer to a comprehensive understanding of the role of DA, its receptors, and cortico-striatal circuits in cognitive flexibility. These studies enhance our knowledge of the neural circuits underlying visual reversal learning and could be relevant for cognitive inflexibility in DA-related disorders, such as Parkinson's disease (Cools et al., 2007), OCD (Denys et al., 2004) or drug-use disorder (Volkow et al., 2009).

Future research should aspire to provide refined detail of the factors involved in behaviour, which could contribute not only to our understanding of the mechanisms, but also to the reproducibility of results. It should also seek to improve the congruence of experimental designs and methods between preclinical and clinical research, and to expose subtle mechanisms that can unveil the bigger picture of how the brain operates to encode and refine complex cognitive processes.

Bibliography

Aarts, E., Maroteaux, G., Loos, M., Koopmans, B., Kovacevic, J., Smit, A. B., Verhage, M., and van der Sluis, S. (2015). The light spot test: Measuring anxiety in mice in an automated home-cage environment. *Behavioural Brain Research*, 294:123–130.

Aizman, O., Brismar, H., Uhlén, P., Zettergren, E., Levey, A. I., Forssberg, H., Greengard, P., and Aperia, A. (2000). Anatomical and physiological evidence for D1 and D2 dopamine receptor colocalization in neostriatal neurons. *Nature Neuroscience*, 3(3):226–230.

Alsiö, J., Nilsson, S. R. O., Gastambide, F., Wang, R. a. H., Dam, S. a., Mar, a. C., Tricklebank, M., and Robbins, T. W. (2015). The role of 5-HT2C receptors in touchscreen visual reversal learning in the rat: a cross-site study. *Psychopharmacology*, 232:4017–4031.

Alsiö, J., Phillips, B. U., Sala-Bayo, J., Nilsson, S. R., Calafat-Pla, T. C., Rizwand, A., Plumbridge, J. M., López-Cruz, L., Dalley, J. W., Cardinal, R. N., Mar, A. C., and Robbins, T. W. (2019). Dopamine D2-like receptor stimulation blocks negative feedback in visual and spatial reversal learning in the rat: behavioural and computational evidence. *Psychopharmacology*, 236:2307–2323.

Amato, D., Milella, M. S., Badiani, A., and Nencini, P. (2006). Compulsive-like effects of repeated administration of quinpirole on drinking behavior in rats. *Behavioural Brain Research*, 172:1–13.

Annett, L. E., McGregor, A., and Robbins, T. W. (1989). The effects of ibotenic acid lesions of the nucleus accumbens on spatial learning and extinction in the rat. *Behavioural Brain Research*, 31(3):231–242.

Anzalone, A., Lizardi-Ortiz, J. E., Ramos, M., De Mei, C., Hopf, F. W., Iaccarino, C., Halbout, B., Jacobsen, J., Kinoshita, C., Welter, M., Caron, M. G., Bonci, A., Sulzer, D., and Borrelli, E. (2012). Dual control of dopamine synthesis and release by presynaptic and postsynaptic dopamine D2 receptors. *The Journal of Neuroscience*, 32(26):9023–34.

Aosaki, T., Graybiel, A. M., and Kimura, M. (1994). Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. *Science*, 265:412–415.

Aquili, L. (2014). The causal role between phasic midbrain dopamine signals and learning. *Frontiers in Behavioral Neuroscience*, 8:2012–2015.

Aquili, L., Liu, A. W., Shindou, M., Shindou, T., and Wickens, J. R. (2014). Behavioral flexibility is increased by optogenetic inhibition of neurons in the nucleus accumbens shell during specific time segments. *Learning and Memory*, 21(4):223–231.

Asaoka, N., Nishitani, N., Kinoshita, H., Nagai, Y., Hatakama, H., Nagayasu, K., Shirakawa, H., Nakagawa, T., and Kaneko, S. (2019). An adenosine A 2A receptor antagonist improves multiple symptoms of repeated quinpirole-induced psychosis. *eNeuro*, 6(1).

Association, A. P. (2013). *Diagnostic and statistical manial of mental disorders: DSM-5*. American Psychiatric Association, 5th ed. edition.

Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *The Journal of Neuroscience*, 27(31):8161–8165.

Balleine, B. W. and Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37:407–419.

Balleine, B. W. and O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, 35(1):48–69.

Bamford, N. S., Robinson, S., Palmiter, R. D., Joyce, J. A., Moore, C., and Meshul, C. K. (2004). Dopamine modulates release from corticostriatal terminals. *The Journal of Neuroscience*, 24(43):9541–9552.

Bari, A., Theobald, D. E., Caprioli, D., Mar, A. C., Aidoo-Micah, A., Dalley, J. W., and Robbins, T. W. (2010). Serotonin Modulates Sensitivity to Reward and Negative Feedback in a Probabilistic Reversal Learning Task in Rats. *Neuropsychopharmacology*, 35(6):1290–1301.

Beats, B. C., Sahakian, B. J., and Levy, R. (1996). Cognitive performance in tests sensitive to frontal lobe dysfunction in the elderly depressed. *Psychological Medicine*, 26(3):591–603.

Beeler, J. A., Daw, N., Frazier, C. R., and Zhuang, X. (2010). Tonic dopamine modulates exploitation of reward learning. *Frontiers in Behavioral Neuroscience*, 4(170):1–14.

Benoit-Marand, M., Borrelli, E., and Gonon, F. (2001). Inhibition of dopamine release via presynaptic D2 receptors: Time course and functional characteristics in vivo. *The Journal of Neuroscience*, 21(23):9134–9141.

Berendse, H. W., Graaf, Y. G., and Groenewegen, H. J. (1992). Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. *The Journal of Comparative Neurology*, 316:314–347.

Berendse, H. W. and Groenewegen, H. J. (1990). Organization of the thalamostriatal projections in the rat, with special emphasis on the ventral striatum. *The Journal of Comparative Neurology*, 299:187–228.

Berridge, K. C. and Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, 28:309–369.

Bertran-Gonzalez, J., Bosch, C., Maroteaux, M., Matamales, M., Hervé, D., Valjent, E., and Girault, J. A. (2008). Opposing patterns of signaling activation in dopamine D1 and D2 receptor-expressing striatal neurons in response to cocaine and haloperidol. *The Journal of Neuroscience*, 28(22):5671–5685.

Besson, M., Belin, D., McNamara, R., Theobald, D. E., Castel, A., Beckett, V. L., Crittenden, B. M., Newman, A. H., Everitt, B. J., Robbins, T. W., and Dalley, J. W. (2010). Dissociable control of impulsivity in rats by dopamine D2/3 receptors in the core and shell subregions of the nucleus accumbens. *Neuropsychopharmacology*, 35(2):560–569.

Birrell, J. M. and Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *The Journal of Neuroscience*, 20(11):4320–4324.

Björklund, A. and Dunnett, S. B. (2007). Dopamine neuron systems in the brain: an update. *Trends in Neurosciences*, 30(5):194–202.

Boeijinga, P. H., Mulder, A. B., Pennartz, C. M. A., Manshanden, I., and Lopes Da Silva, F. H. (1993). Responses of the nucleus accumbens following fornix/fimbria stimulation in the rat. Identification and long-term potentiation of mono- and polysynaptic pathways. *Neuroscience*, 53(4):1049–1058.

Boekhoudt, L., Wijbrans, E. C., Man, J. H., Luijendijk, M. C., de Jong, J. W., van der Plasse, G., Vanderschuren, L. J., and Adan, R. A. (2018). Enhancing excitability of dopamine neurons promotes motivational behaviour through increased action initiation. *European Neuropsychopharmacology*, 28(1):171–184.

Borea, P. A., Gessi, S., Merighi, S., Vincenzi, F., and Varani, K. (2018). Pharmacology of adenosine receptors: The state of the art. *Physiological Reviews*, 98(3):1591–1625.

Borek, L. L., Amick, M. M., and Friedman, J. H. (2006). Non-motor aspects of Parkinson's disease. *CNS Spectrum*, 11(7):541–554.

Boulougouris, V., Castañé, A., and Robbins, T. W. (2009). Dopamine D2/D3 receptor agonist quinpirole impairs spatial reversal learning in rats: Investigation of D3 receptor involvement in persistent behavior. *Psychopharmacology*, 202:611–620.

Boulougouris, V., Dalley, J. W., and Robbins, T. W. (2007). Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. *Behavioural Brain Research*, 179(2):219–228.

Boulougouris, V., Glennon, J. C., and Robbins, T. W. (2008). Dissociable effects of selective 5-HT2A and 5-HT2C receptor antagonists on serial spatial reversal learning in rats. *Neuropsychopharmacology*, 33:2007–2019.

Boulougouris, V. and Robbins, T. W. (2010). Enhancement of Spatial Reversal Learning by 5-HT2C Receptor Antagonism Is Neuroanatomically Specific. *The Journal of Neuroscience*, 30(3):930–938.

Boyson, S. J., McGonigle, P., and Molinoff, P. B. (1986). Quantitative autoradiographic localization of the D1 and D2 subtypes of dopamine receptors in rat brain. *The Journal of Neuroscience*, 6(11):3177–3188.

Brigman, J. L., Daut, R. a., Wright, T., Gunduz-Cinar, O., Graybeal, C., Davis, M. I., Jiang, Z., Saksida, L. M., Jinde, S., Pease, M., Bussey, T. J., Lovinger, D. M., Nakazawa, K., and Holmes, A. (2013). GluN2B in corticostriatal circuits governs choice learning and choice shifting. *Nature neuroscience*, 16(8):1101–10.

Brown, M. T., Tan, K. R., O'Connor, E. C., Nikonenko, I., Muller, D., and Lüscher, C. (2012). Ventral tegmental area GABA projections pause accumbal cholinergic interneurons to enhance associative learning. *Nature*, 492(7429):452–456.

Burke, D. A., Roitstein, H. G., and Alvarez, V. A. (2017). Striatal local circuitry: a new framework for lateral inhibition. *Neuron*, 96:267–284.

Cachope, R., Mateo, Y., Mathur, B. N., Irving, J., Wang, H. L., Morales, M., Lovinger, D. M., and Cheer, J. F. (2012). Selective activation of cholinergic interneurons enhances accumbal phasic dopamine release: Setting the tone for reward processing. *Cell Reports*, 2(1):33–41.

Cagniard, B., Beeler, J. A., Britt, J. P., McGehee, D. S., Marinelli, M., and Zhuang, X. (2006). Dopamine scales performance in the absence of new learning. *Neuron*, 51(5):541–547.

Calabresi, P., Picconi, B., Tozzi, A., and Di Filippo, M. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends in Neurosciences*, 30(5):211–219.

Cardinal, R. N. and Everitt, B. J. (2004). Neural and psychological mechanisms underlying appetitive learning: link to drug addiction. *Current Opinion in Neurobiology*, 14:156–162.

Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, 26(3):321–352.

Carli, M., Robbins, T. W., Evenden, J. L., and Everitt, B. J. (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.

Carlsson, A., Lindgvist, M., Magnuson, T., and Waldeck, B. (1958). On the presence of 3-hydroxytyramine in brain. *Science*, 127:471.

Castañé, A., Theobald, D. E. H., and Robbins, T. W. (2010). Selective lesions of the dorsomedial striatum impair serial spatial reversal learning in rats. *Behavioural Brain Research*, 210(1):74–83.

Castro, D. C. and Bruchas, M. R. (2019). A motivational and neuropeptidergic hub: Anatomical and functional fiversity within the nucleus accumbens shell. *Neuron*, 102(3):529–552.

Cazorla, M., de Carvalho, F. D., Chohan, M. O., Shegda, M., Chuhma, N., Rayport, S., Ahmari, S. E., Moore, H., and Kellendonk, C. (2014). Dopamine D2 receptors regulate the anatomical and functional balance of basal ganglia circuitry. *Neuron*, 81(1):153–64.

Centonze, D., Gubellini, P., Usiello, A., Rossi, S., Tscherter, A., Bracci, E., Erbs, E., Tognazzi, N., Bernardi, G., Pisani, A., Calabresi, P., and Borrelli, E. (2004). Differential contribution of dopamine D2S and D2L receptors in the modulation of glutamate and GABA transmission in the striatum. *Neuroscience*, 129(1):157–166.

Centonze, D., Picconi, B., Baunez, C., Borrelli, E., Pisani, A., Bernardi, G., and Calabresi, P. (2002). Cocaine and amphetamine depress striatal GABAergic synaptic transmission through D2 dopamine receptors. *Neuropsychopharmacology*, 26(2):164–175.

Chamberlain, S. R., Fineberg, N. A., Menzies, L. A., Blackwell, A. D., Bullmore, E. T., Robbins, T. W., and Sahakian, B. J. (2007). Impaired cognitive flexibility and motor inhibition in unaffected first-degree relatives of patients with obsessive-compulsive disorder. *American Journal of Psychiatry*, 164(2):335–338.

Chang, C. Y., Esber, G. R., Marrero-Garcia, Y., Yau, H. J., Bonci, A., and Schoenbaum, G. (2015). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nature Neuroscience*, 19(1):111–116.

Clark, L., Chamberlain, S. R., and Sahakian, B. J. (2009). Neurocognitive mechanisms in depression: Implications for treatment. *Annual Review of Neuroscience*, 32(1):57–74.

Clarke, H. F., Cardinal, R. N., Rygula, R., Hong, Y. T., Fryer, T. D., Sawiak, S. J., Ferrari, V., Cockcroft, G., Aigbirhio, F. I., Robbins, T. W., and Roberts, A. C. (2014). Orbitofrontal dopamine depletion upregulates caudate dopamine and alters behavior via changes in reinforcement sensitivity. *The Journal of Neuroscience*, 34(22):7663–76.

Clarke, H. F., Dalley, J. W., Crofts, H. S., Robbins, T. W., and Roberts, A. C. (2004). Cognitive inflexibility after prefrontal serotonin depletion. *Science*, 304(5672):878–880.

Clarke, H. F., Hill, G. J., Robbins, T. W., and Roberts, A. C. (2011). Dopamine, but not serotonin, regulates reversal learning in the marmoset caudate nucleus. *The Journal of Neuroscience*, 31(11):4290–4297.

Clarke, H. F., Robbins, T. W., and Roberts, A. C. (2008). Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *The Journal of Neuroscience*, 28(43):10972–10982.

Clarke, H. F., Walker, S. C., Dalley, J. W., Robbins, T. W., and Roberts, A. C. (2007). Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. *Cerebral Cortex*, 17(1):18–27.

Clatworthy, P. L., Lewis, S. J. G., Brichard, L., Hong, Y. T., Izquierdo, D., Clark, L., Cools, R., Aigbirhio, F. I., Baron, J.-C., Fryer, T. D., and Robbins, T. W. (2009). Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory. *The Journal of Neuroscience*, 29(15):4690–4696.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers., 2nd ed. edition.

Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*, 482(7383):85–88.

Cole, S. L., Robinson, M. J., and Berridge, K. C. (2018). Optogenetic self-stimulation in the nucleus accumbens: D1 reward versus D2 ambivalence. *PLoS ONE*, 13(11):1–29.

Coney, A. M. and Marshall, J. M. (1998). Role of adenosine and its receptors in the vasodilatation induced in the cerebral cortex of the rat by systemic hypoxia. *Journal of Physiology*, 509(2):507–518.

Cools, A. R. and Van Rossum, J. M. (1976). Excitation-mediating and inhibition-mediating dopamine-receptors: A new concept towards a better understanding of electrophysiological, biochemical, pharmacological, functional and clinical data. *Psychopharmacologia*, 45(3):243–254.

Cools, R., Barker, R. A., Sahakian, B. J., and Robbins, T. W. (2001). Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cerebral cortex*, 11(12):1136–1143.

Cools, R., Clark, L., Owen, A. M., and Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *The Journal of Neuroscience*, 22(11):4563–4567.

Cools, R., Lewis, S. J. G., Clark, L., Barker, R. A., and Robbins, T. W. (2007). L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease. *Neuropsychopharmacology*, 32(1):180–189.

Corbit, L. H., Muir, J. L., and Balleine, B. W. (2001). The role of the nucleus accumbens in instrumental conditioning: evidence of a functional dissociation between accumbens core and shell. *The Journal of Neuroscience*, 21(9):3251–3260.

Corbit, L. H., Nie, H., and Janak, P. H. (2014). Habitual responding for alcohol depends upon both AMPA and D2 receptor signaling in the dorsolateral striatum. *Frontiers in Behavioral Neuroscience*, 8(September):1–9.

Costantini, A. F. and Hoving, K. L. (1973). The relationship of cognitive and motor response inhibition to age and IQ. *Journal of Genetic Psychology*, 123:309–319.

Cox, J. and Witten, I. B. (2019). Striatal circuits for reward learning and decision-making. *Nature Reviews Neuroscience*, 20(8).

Cox, S. M., Frank, M. J., Larcher, K., Fellows, L. K., Clark, C. A., Leyton, M., and Dagher, A. (2015). Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. *NeuroImage*, 109:95–101.

Culver, K. E., Szechtman, H., and Levant, B. (2008). Altered dopamine D2-like receptor binding in rats with behavioral sensitization to quinpirole: effects of pre-treatment with Ro 41-1049. *European Journal of Pharmacology*, 592:67–72.

Dagher, A. and Robbins, T. W. (2009). Personality, addiction, dopamine: insights from Parkinson's disease. *Neuron*, 61(4):502–510.

Dahlström, A. and Fuxe, K. (1964). Localization of monoamines in the lower brain stem. *Experientia*, 20(7):398–399.

Dajani, D. R. and Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, 38(9):571–578.

Dalley, J. W., Cardinal, R. N., and Robbins, T. W. (2004). Prefrontal executive and cognitive functions in rodents: Neural and neurochemical substrates. *Neuroscience and Biobehavioral Reviews*, 28(7):771–784.

Dalley, J. W. and Robbins, T. W. (2017). Fractionating impulsivity: Neuropsychiatric implications. *Nature Reviews Neuroscience*, 18(3):158–171.

Dalton, G. L., Phillips, A. G., and Floresco, S. B. (2014). Preferential involvement by nucleus accumbens shell in mediating probabilistic learning and reversal shifts. *The Journal of Neuroscience*, 34(13):4618–4626.

Dalton, G. L., Wang, N. Y., Phillips, A. G., and Floresco, S. B. (2016). Multifaceted contributions by different regions of the orbitofrontal and medial prefrontal cortex to probabilistic reversal learning. *The Journal of Neuroscience*, 36(6):1996–2006.

Daly, J. W., Butts-Lamb, P., and Padgett, W. (1983). Subclasses of adenosine receptors in the central nervous system: Interaction with caffeine and related methylxanthines. *Cellular and Molecular Neurobiology*, 3(1):69–80.

Dauer, W. and Przedborski, S. (2003). Review Parkinson's disease : Mechanisms and models. *Neuron*, 39:889–909.

Daw, N. D. (2009). Trial-by-trial data analysis using computational models. In *Decision Making, Affect, and Learning: Attention and Performance XXIII*, pages 1–26. OUP Oxford.

Day, M., Wang, Z., Ding, J., An, X., Ingham, C. a., Shering, A. F., Wokosin, D., Ilijic, E., Sun, Z., Sampson, A. R., Mugnaini, E., Deutch, A. Y., Sesack, S. R., Arbuthnott, G. W., and Surmeier, D. J. (2006). Selective elimination of glutamatergic synapses on striatopallidal neurons in Parkinson disease models. *Nature Neuroscience*, 9(2):251–259.

Day, M., Wokosin, D., Plotkin, J. L., Tian, X., and Surmeier, D. J. (2008). Differential excitability and modulation of striatal medium spiny neuron dendrites. *The Journal of Neuroscience*, 28(45):11603–11614.

de Haas, R., Nijdam, A., Westra, T. A., Kas, M. J. H., and Westenberg, H. G. M. (2011). Behavioral pattern analysis and dopamine release in quinpirole-induced repetitive behavior in rats. *Journal of Psychopharmacology*, 25(12):1712–1719.

De Keyser, J., Claeys, A., De Backer, J.-P., Ebinger, G., Roels, F., and Vauquelin, G. (1988). Autoradiographic localization of D1 and D2 dopamine binding sites in the human retina. *Neuroscience Letters*, 91:142–147.

De Mei, C., Ramos, M., Iitaka, C., and Borrelli, E. (2009). Getting specialized: presynaptic and postsynaptic dopamine D2 receptors. *Current Opinion in Pharmacology*, 9(1):53–58.

Deisseroth, K. (2011). Optogenetics. *Nature Methods*, 8(1):26–29.

Delle Donne, K. T., Sesack, S. R., and Pickel, V. M. (1996). Ultrastructural immunocytochemical localization of neurotensin and the dopamine D2 receptor in the rat nucleus accumbens. *Journal of Comparative Neurology*, 371(4):552–566.

DeMet, E. M. and Chicz-DeMet, A. (2002). Localization of adenosine A2A-receptors in rat brain with [3H]ZM-241385. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 366(5):478–481.

den Ouden, H. E. M., Daw, N. D., Fernandez, G., Elshout, J. A., Rijpkema, M., Hoogman, M., Franke, B., and Cools, R. (2013). Dissociable effects of dopamine and serotonin on reversal learning. *Neuron*, 80(4):1090–100.

Denys, D., van der Wee, N., Janssen, J., De Geus, F., and Westenberg, H. G. M. (2004). Low level of dopaminergic D2 receptor binding in obsessive-compulsive disorder. *Biological psychiatry*, 55(10):1041–5.

Deutch, A. Y. and Cameron, D. S. (1992). Pharmacological characterization of dopamine systems in the nucleus accumbens core and shell. *Neuroscience*, 46(1):49–56.

Dhawan, S. S., Tait, D. S., and Brown, V. J. (2019). More rapid reversal learning following overtraining in the rat is evidence that behavioural and cognitive flexibility are dissociable. *Behavioural Brain Research*, 363:45–52.

Dixon, A. K., Gubitz, A. K., Sirinathsinghji, D. J., Richardson, P. J., and Freeman, T. C. (1996). Tissue distribution of adenosine receptor mRNAs in the rat. *British Journal of Pharmacology*, 118(6):1461–1468.

Dumartin, B., Doudnikoff, E., Gonon, F., and Bloch, B. (2007). Differences in ultrastructural localization of dopaminergic D1 receptors between dorsal striatum and nucleus accumbens in the rat. *Neuroscience Letters*, 419:273–277.

Eagle, D. M., Bari, A., and Robbins, T. W. (2008). The neuropsychopharmacology of action inhibition: Cross-species translation of the stop-signal and go/no-go tasks. *Psychopharmacology*, 199(3):439–456.

Eagle, D. M., Schepisi, C., Chugh, S., Desai, S., Han, S. Y. S., Huang, T., Lee, J. J., Sobala, C., Ye, W., Milton, A. L., and Robbins, T. W. (2020). Dissociable dopaminergic and pavlovian influences in goal-trackers and sign-trackers on a model of compulsive checking OCD. *Psychopharmacology*, pages 1–13.

Economidou, D., Theobald, D. E. H., Robbins, T. W., Everitt, B. J., and Dalley, J. W. (2012). Norepinephrine and dopamine modulate impulsivity on the five-choice serial reaction time task through opponent actions in the shell and core sub-regions of the nucleus accumbens. *Neuropsychopharmacology*, 37(9):2057–2066.

Eilam, D. and Szechtman, H. (1989). Biphasic effect of D-2 agonist quinpirole on locomotion and movements. *European Journal of Pharmacology*, 161(2-3):151–157.

Elliott, R., Sahakian, B. J., Herrod, J. J., Robbins, T. W., and Paykel, E. S. (1997). Abnormal response to negative feedback in unipolar depression: Evidence for a diagnosis specific impairment. *Journal of Neurology Neurosurgery and Psychiatry*, 63(1):74–82.

Epstein, J., Pan, H., Kocsis, J. H., Yang, Y., Butler, T., Chusid, J., Hochberg, H., Murrough, J., Strohmayer, E., Stern, E., and Silbersweig, D. A. (2006). Lack of ventral striatal response to positive stimuli in depressed versus normal subjects. *American Journal of Psychiatry*, 163(10):1784–1790.

Ersche, K. D., Roiser, J. P., Abbott, S., Craig, K. J., Mller, U., Suckling, J., Ooi, C., Shabbir, S. S., Clark, L., Sahakian, B. J., Fineberg, N. A., Merlo-Pich, E. V., Robbins, T. W., and Bullmore, E. T. (2011). Response perseveration in stimulant dependence is associated with striatal dysfunction and can be ameliorated by a D2/3receptor agonist. *Biological Psychiatry*, 70(8):754–762.

Escobar, A. P., González, M. P., Meza, R. C., Noches, V., Henny, P., Gysling, K., España, R. A., Fuentealba, J. A., and Andrés, M. E. (2017). Mechanisms of kappa opioid receptor potentiation of dopamine D2 receptor function in quinpirole- Induced locomotor sensitization in rats. *International Journal of Neuropsychopharmacology*, 20(8):660–669.

Eslinger, P. J. and Grattan, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia*, 31(1):17–28.

Evenden, J. L. and Robbins, T. W. (1985). The effects of d-amphetamine, chlordiazepoxide and alpha-flupenthixol on food-reinforced tracking of a visual stimulus by rats. *Psychopharmacology*, 85(3):361–366.

Faure, A., Haberland, U., Condé, F., and El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *The Journal of Neuroscience*, 25(11):2771–2780.

Ferré, S. (2008). An update on the mechanisms of the psychostimulant effects of caffeine. *Journal of Neurochemistry*, 105(4):1067–1079.

Fields, H. L., Hjelmstad, G. O., Margolis, E. B., and Nicola, S. M. (2007). Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement. *Annual Review of Neuroscience*, 30(1):289–316.

Fineberg, N. A., Potenza, M. N., Chamberlain, S. R., Berlin, H. A., Menzies, L., Bechara, A., Sahakian, B. J., Robbins, T. W., Bullmore, E. T., and Hollander, E. (2010). Probing compulsive and impulsive behaviors, from animal models to endophenotypes: A narrative review. *Neuropsychopharmacology*, 35:591–604.

Flagel, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czuj, A., Willuhn, I., Akers, C. A., Clinton, S. M., Phillips, P. E., and Akil, H. (2011). A selective role for dopamine in stimulus-reward learning. *Nature*, 469(7328):53–59.

Flint, J. and Munafò, M. R. (2007). The endophenotype concept in psychiatric genetics. *Psychological Medicine*, 37(2):163–180.

Floresco, S. B., Ghods-Sharifi, S., Vexelman, C., and Magyar, O. (2006). Dissociable roles for the nucleus accumbens core and shell in regulating set shifting. *The Journal of Neuroscience*, 26(9):2449–2457.

Floresco, S. B., West, A. R., Ash, B., Moorel, H., and Grace, A. A. (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nature Neuroscience*, 6(9):968–973.

Floresco, S. B., Zhang, Y., and Enomoto, T. (2009). Neural circuits subserving behavioral flexibility and their relevance to schizophrenia. *Behavioural Brain Research*, 204(2):396–409.

Ford, C. P. (2014). The role of D2-autoreceptors in regulating dopamine neuron activity and transmission. *Neuroscience*, 282:13–22.

Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, 306(5703):1940–1943.

Fraser, K. M., Haight, J. L., Gardner, E. L., and Flagel, S. B. (2016). Examining the role of dopamine D2 and D3 receptors in Pavlovian conditioned approach behaviors. *Behavioural Brain Research*, 305:87–99.

Freed, C. R. and Yamamoto, B. K. (1985). Regional brain dopamine metabolism: A marker for the speed, direction, and posture of moving animals. *Science*, 229:62–65.

Frozi, J., de Carvalho, H. W., Ottoni, G. L., Cunha, R. A., and Lara, D. R. (2018). Distinct sensitivity to caffeine-induced insomnia related to age. *Journal of Psychopharmacology*, 32(1):89–95.

Furlong, T. M., Supit, A. S., Corbit, L. H., Killcross, S., and Balleine, B. W. (2017). Pulling habits out of rats: adenosine 2A receptor antagonism in dorsomedial striatum rescues meth-amphetamine-induced deficits in goal-directed action. *Addiction Biology*, 22(1):172–183.

Futami, T., Takakusaki, K., and Kitai, S. T. (1995). Glutamatergic and cholinergic inputs from the pedunculopontine tegmental nucleus to dopamine neurons in the substantia nigra pars compacta. *Neuroscience Research*, 21(4):331–342.

Galsworthy, M. J., Amrein, I., Kupstov, P. A., Poletaeva, I. I., Zinn, P., Rau, A., Vyssotski, A., and Lipp, H.-P. (2005). A comparison of wild-caught wood mice and bank voles in the Intellicage: assessing exploration, daily activity patterns and place learning paradigms. *Behavioural Brain Research*, 157:211–217.

Gantz, S. C., Robinson, B. G., Buck, D. C., Bunzow, J. R., Neve, R. L., Williams, J. T., and Neve, K. A. (2015). Distinct regulation of dopamine D2S and D2L autoreceptor signaling by calcium. *eLife*, 4(e09358):1–19.

Garner, J. P., Thogerson, C. M., Würbel, H., Murray, J. D., and Mench, J. A. (2006). Animal neuropsychology: Validation of the intra-dimensional extra-dimensional set shifting task for mice. *Behavioural Brain Research*, 173(1):53–61.

Gerfen, C. P. (2004). *Basal ganglia. In: The rat nervous system*. San Diego: Elsevier, 3rd edition.

Gerfen, C. R., Engber, T. M., Mahan, L. C., Susel, Z., Chase, T. N., Monsma, F. J., and Sibley, D. R. (1990). D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science*, 250(4986):1429–1432.

Gertler, T. S., Chan, C. S., and Surmeier, D. J. (2008). Dichotomous anatomical properties of adult striatal medium spiny neurons. *The Journal of Neuroscience*, 28(43):10814–10824.

Gillan, C. M., Robbins, T. W., Sahakian, B. J., van den Heuvel, O. A., and van Wingen, G. (2016). The role of habit in compulsivoty. *European Neuropsychopharmacology*, 26:828–840.

Giros, B., Sokoloff, P., Martres, M. P., Riou, J. F., Emorine, L. J., and Schwartz, J.-C. (1989). Alternative splicing directs the expression of two D2 dopamine receptor isoforms. *Nature*, 342:923–926.

Godier, L. R. and Park, R. J. (2014). Compulsivity in anorexia nervosa: A transdiagnostic concept. *Frontiers in Psychology*, 5(778):1–18.

Goldsmith, D. R., Haroon, E., Woolwine, B. J., Jung, M. Y., Wommack, E. C., Harvey, P. D., Treadway, M. T., Felger, J. C., and Miller, A. H. (2016). Inflammatory markers are associated with decreased psychomotor speed in patients with major depressive disorder. *Brain, Behavior, and Immunity*, 56:281–288.

Goto, Y. and Grace, A. A. (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nature Neuroscience*, 8(6):805–812.

Goto, Y., Otani, S., and Grace, A. A. (2007). The Yin and Yang of dopamine release: a new perspective. *Neuropharmacology*, 53(5):583–587.

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

Grace, A. A. (2000). The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction*, 95(8):119–128.

Grace, A. A. and Bunney, B. S. (1979). Paradoxical GABA excitation of nigral dopaminergic cells: Indirect mediation through reticulata inhibitory neurons. *European Journal of Pharmacology*, 59(3-4):211–218.

Grace, A. A. and Bunney, B. S. (1984). The control of firing pattern in nigral dopamine neurons: Burst firing. *The Journal of Neuroscience*, 4(11):2877–2890.

Grace, A. A. and Onn, S. P. (1989). Morphology and electrophysiological properties of immunocytochemistry identified rat dopamine neurons recorded in vitro. *The Journal of Neuroscience*, 9(10):3463–3481.

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., Reid, I., Hall, J., and Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*, 134(6):1751–1764.

Graybeal, C., Feyder, M., Schulman, E., Saksida, L. M., Bussey, T. J., Brigman, J. L., and Holmes, A. (2011). Paradoxical reversal learning enhancement by stress or prefrontal cortical damage: rescue with BDNF. *Nature Neuroscience*, 14(12):1507–1509.

Gremel, C. M. and Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature communications*, 4(2264):1–12.

Groenewegen, H. J., Wright, C. I., Beijer, A. V., and Voorn, P. (1999). Convergence and segregation of ventral striatal inputs and outputs. *Annals of the New York Academy of Sciences*, 877:49–63.

Groman, S. M., James, A. S., Seu, E., Tran, S., Clark, T. A., Harpster, S. N., Crawford, M., Burtner, J. L., Feiler, K., Roth, R. H., Elsworth, J. D., London, E. D., and Jentsch, J. D. (2014). In the blink of an eye: Relating positive-feedback sensitivity to striatal dopamine D2-like receptors through blink rate. *The Journal of Neuroscience*, 34(43):14443–14454.

Groman, S. M., Keistler, C., Keip, A. J., Hammarlund, E., DiLeone, R. J., Pittenger, C., Lee, D., and Taylor, J. R. (2019). Orbitofrontal circuits control multiple reinforcement-learning processes. *Neuron*, 103(4):734–746.e3.

Groman, S. M., Lee, B., London, E. D., Mandelkern, M. A., James, A. S., Feiler, K., Rivera, R., Dahlbom, M., Sossi, V., Vandervoort, E., and Jentsch, J. D. (2011). Dorsal striatal D2-like receptor availability covaries with sensitivity to positive reinforcement during discrimination learning. *The Journal of Neuroscience*, 31(20):7291–7299.

Gruber, A. J. and McDonald, R. J. (2012). Context, emotion, and the strategic pursuit of goals: Interactions among multiple brain systems controlling motivated behavior. *Frontiers in Behavioral Neuroscience*, 6(50):1–26.

Gruner, P. and Pittenger, C. (2017). Cognitive inflexibility in Obsessive-Compulsive Disorder. *Neuroscience*, 345:243–255.

Gu, B. M., Park, J. Y., Kang, D. H., Lee, S. J., Yoo, S. Y., Jo, H. J., Choi, C. H., Lee, J. M., and Kwon, J. S. (2008). Neural correlates of cognitive inflexibility during task-switching in obsessive-compulsive disorder. *Brain*, 131:155–164.

Guzmán, J. N., Hernández, A., Galarraga, E., Tapia, D., Laville, A., Vergara, R., Aceves, J., and Bargas, J. (2003). Dopaminergic modulation of axon collaterals interconnecting spiny neurons of the rat striatum. *The Journal of Neuroscience*, 23(26):8931–8940.

Haber, S. N. (2016). Corticostriatal circuitry. *Dialogues in Clinical Neuroscience*, 18(1):7–21.

Hales, C. A., Stuart, S. A., Anderson, M. H., and Robinson, E. S. J. (2014). Modelling cognitive affective bias in major depressive disorder using rodents. *British Journal of Pharmacology*, 171:4524–4538.

Haluk, D. M. and Floresco, S. B. (2009). Ventral striatal dopamine modulation of different forms of behavioral flexibility. *Neuropsychopharmacology*, 34:2041–2052.

Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M., Kennedy, R. T., Aragona, B. J., and Berke, J. D. (2016). Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, 19(1):117–126.

Hart, A. S., Rutledge, R. B., Glimcher, P. W., and Phillips, P. E. M. (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *The Journal of Neuroscience*, 34(3):698–704.

Hassani, O. K., François, C., Yelnik, J., and Féger, J. (1997). Evidence for a dopaminergic innervation of the subthalamic nucleus in the rat. *Brain Research*, 749(1):88–94.

Hervig, M. E., Fiddian, L., Piilgaard, L., Božič, T., Blanco-Pozo, M., Knudsen, C., Olesen, S. F., Alsiö, J., and Robbins, T. W. (2019). Dissociable and paradoxical roles of rat medial and lateral orbitofrontal cortex in visual serial reversal learning. *Cerebral Cortex*, 00:1–14.

Hikosaka, K. and Watanabe, M. (2000). Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cerebral Cortex*, 10(3):263–271.

Hollerman, J. R. and Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1(4):304–309.

Horner, A. E., Heath, C. J., Hvoslef-Eide, M., Kent, B. A., Kim, C. H., Nilsson, S. R. O., Alsiö, J., Oomen, C. A., Holmes, A., Saksida, L. M., and Bussey, T. J. (2013). The touchscreen operant platform for testing learning and memory in rats and mice. *Nature protocols*, 8(10):1961–84.

Horst, N. K., Jupp, B., Roberts, A. C., and Robbins, T. W. (2019). D2 receptors and cognitive flexibility in marmosets: tri-phasic dose–response effects of intra-striatal quinpirole on serial reversal performance. *Neuropsychopharmacology*, 44(3):564–571.

Howe, M. W. and Dombeck, D. A. (2016). Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature*, 535(7613):505–510.

Howe, M. W., Tierney, P. L., Sandberg, S. G., Phillips, P. E., and Graybiel, A. M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature*, 500(7464):575–579.

Hsu, K.-S., Huang, C.-C., Yang, C.-H., and Gean, P.-W. (1995). Presynaptic D2 dopaminergic receptors mediate inhibition of excitatory synmaptic transmission in rat neos. *Brain Research*, 690:264–268.

Humphries, M. D. and Prescott, T. J. (2010). The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. *Progress in Neurobiology*, 90(4):385–417.

Hung, A. Y. and Schwarzschild, M. A. (2014). Treatment of Parkinson's disease: What's in the non-dopaminergic pipeline? *Neurotherapeutics*, 11(1):34–46.

Ikemoto, S. (2007). Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Research Reviews*, 56(1):27–78.

Ineichen, C., Sigrist, H., Spinelli, S., Lesch, K. P., Sautter, E., Seifritz, E., and Pryce, C. R. (2012). Establishing a probabilistic reversal learning test in mice: Evidence for the processes mediating reward-stay and punishment-shift behaviour and for their modulation by serotonin. *Neuropharmacology*, 63(6):1012–1021.

Iordanova, M. D., Westbrook, R. F., and Killcross, A. S. (2006). Dopamine activity in the nucleus accumbens modulates blocking in fear conditioning. *European Journal of Neuroscience*, 24(11):3265–3270.

Isomura, Y., Takekawa, T., Harukuni, R., Handa, T., Aizawa, H., Takada, M., and Fukai, T. (2013). Reward-modulated motor information in identified striatum neurons. *The Journal of Neuroscience*, 33(25):10209–10220.

Iversen, S. D. and Iversen, L. L. (2007). Dopamine: 50 years in perspective. *Trends in Neurosciences*, 30(5):188–193.

Izquierdo, A. (2017). Functional heterogeneity within rat orbitofrontal cortex in reward learning and decision making. *The Journal of Neuroscience*, 37(44):10529.

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., and Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, 345:12–26.

Izquierdo, A. and Jentsch, J. D. (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology*, 219(2):607–620.

Izquierdo, A., Wiedholz, L. M., Millstein, R. A., Yang, R. J., Bussey, T. J., Saksida, L. M., and Holmes, A. (2006). Genetic and dopaminergic modulation of reversal learning in a touchscreen-based operant procedure for mice. *Behavioural Brain Research*, 171:181–188.

Jackson, D. M. and Westlind-Danielsson, A. (1994). Dopamine receptors: Molecular biology, biochemistry and behavioural aspects. *Pharmacology and Therapeutics*, 64(2):291–370.

Jackson, S. A., Horst, N. K., Axelsson, S. F., Horiguchi, N., Cockcroft, G. J., Robbins, T. W., and Roberts, A. C. (2019). Selective role of the putamen in serial reversal learning in the marmoset. *Cerebral Cortex*, 29(1):447–460.

Jaquins-Gerstl, A. and Michael, A. C. (2015). A review of the effects of FSCV and microdialysis measurements on dopamine release in the surrounding tissue. *Analyst*, 140(11):3696–3708.

Jocham, G., Klein, T. A., Neumann, J., Von Cramon, D. Y., Reuter, M., and Ullsperger, M. (2009). Dopamine DRD2 polymorphism alters reversal learning and associated neural activity. *The Journal of Neuroscience*, 29(12):3695–3704.

Jones, B. and Mishkin, M. (1972). Limbic lesions and the problem of stimulus-reinforcement associations. *Experimental Neurology*, 36(2):362–377.

Kanen, J. W., Ersche, K. D., Fineberg, N. A., Robbins, T. W., and Cardinal, R. N. (2019). Computational modelling reveals contrasting effects on reinforcement learning and cognitive flexibility in stimulant use disorder and obsessive-compulsive disorder: remediating effects of dopaminergic D2/3 receptor agents. *Psychopharmacology*, 236(8):2337–2358.

Keedwell, P. A., Andrew, C., Williams, S. C., Brammer, M. J., and Phillips, M. L. (2005). The neural correlates of anhedonia in major depressive disorder. *Biological Psychiatry*, 58(11):843–853.

Keefe, K. A., Zigmod, M. J., and Abercrombie, E. D. (1993). In vivo regulation of extracellular dopamine in the neostriatum: influence of impulse activity and local excitatory amino acids. *Journal of Neural Transmission*, 91:223–240.

Kehagia, A., Murray, G. K., and Robbins, T. W. (2010). Learning and cognitive flexibility: Frontostriatal function and monoaminergic modulation. *Current Opinion in Neurobiology*, 20(2):169–192.

Keiflin, R., Pribut, H. J., Shah, N. B., and Janak, P. H. (2019). Ventral tegmental dopamine neurons participate in reward identity predictions. *Current Biology*, 29(1):93–103.

Keistler, C., Barker, J. M., and Taylor, J. R. (2015). Infralimbic prefrontal cortex interacts with nucleus accumbens shell to unmask expression of outcomeselective Pavlovian-to-instrumental transfer. *Learning and Memory*, 22(10):509–513.

Kim, H., Malik, A., Mikhael, J., Bech, P., Tsutsui-Kimura, I., Sun, F., Zhang, Y., Li, Y., Watabe-Uchida, M., Gershman, S., and Uchida, N. (2019). A unified framework for dopamine signals across timescales. *bioRxiv*.

Klanker, M., Feenstra, M., and Denys, D. (2013). Dopaminergic control of cognitive flexibility in humans and animals. *Frontiers in Neuroscience*, 7(7 NOV):1–24.

Klanker, M., Fellinger, L., Feenstra, M., Willuhn, I., and Denys, D. (2017). Regionally distinct phasic dopamine release patterns in the striatum during reversal learning. *Neuroscience*, 345(5):110–123.

Klanker, M., Sandberg, T., Joosten, R., Willuhn, I., Feenstra, M., and Denys, D. (2015). Phasic dopamine release induced by positive feedback predicts individual differences in reversal learning. *Neurobiology of Learning and Memory*, 125:135–145.

Knight, A. R., Misra, A., Quirk, K., Benwell, K., Revell, D., Kennett, G., and Bickerdike, M. (2004). Pharmacological characterisation of the agonist radioligand binding site of 5-HT2A, 5-HT2B and 5-HT2C receptors. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 370(2):114–123.

Kruzich, P. J. and Grandy, D. K. (2004). Dopamine D2 receptors mediate two-odor discrimination and reversal learning in C57BL/6 mice. *BMC neuroscience*, 5(12):1–10.

Kupchik, Y. M., Brown, R. M., Heinsbroek, J. A., Lobo, M. K., Schwartz, D. J., and Kalivas, P. W. (2015). Coding the direct/indirect pathways by D1 and D2 receptors is not valid for accumbens projections. *Nature Neuroscience*, 18(9):1230–1232.

Lak, A., Stauffer, W. R., and Schultz, W. (2014). Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6):2343–2348.

Lanciego, J. L., Luquin, N., and Obeso, J. A. (2012). Functional neuroanatomy of the basal ganglia. *Cold Spring Harbor Perspectives in Medicine*, 2(a009621).

Laughlin, R. E., Grant, T. L., Williams, R. W., and Jentsch, J. D. (2011). Genetic dissection of behavioral flexibility : Reversal learning in mice. *Biological Psychiatry*, 69:1109–1116.

Lauwereyns, J., Watanabe, K., Coe, B., and Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature*, 418:413–417.

Lavy, E., Van Oppen, P., and Van Den Hout, M. (1994). Selective processing of emotional information in obsessive compulsive disorder. *Behaviour Research and Therapy*, 32(2):243–246.

Le Moine, C. and Bloch, B. (1996). Expression of the D3 dopamine receptor in peptidergic neurons of the nucleus accumbens: comparison with the D1 and D2 dopamine receptors. *Neuroscience*, 73(1):131–143.

Lee, B., Groman, S., London, E. D., and Jentsch, J. D. (2007). Dopamine D2/D3 receptors play a specific role in the reversal of a learned visual discrimination in monkeys. *Neuropsychopharmacology*, 32(10):2125–2134.

Leeson, V. C., Robbins, T. W., Matheson, E., Hutton, S. B., Ron, M. A., Barnes, T. R. E., and Joyce, E. M. (2009). Discrimination learning, reversal, and set-shifting in first-episode schizophrenia: Stability over six years and specific associations with medication type and disorganization syndrome. *Biological Psychiatry*, 66:586–593.

Leibenluft, E., Fiero, P. L., Bartko, J. J., Moul, D. E., and Rosenthal, N. E. (1993). Depressive symptoms and the self-reported use of alcohol, caffeinve, and carbohydrates in normal volunteers and four groups of psychiatric outpatients. *American Journal of Psychiatry*, 150(2):294–301.

Li, X., Qi, J., Yamaguchi, T., Wang, H.-L., and Morales, M. (2012). Heterogeneous composition of dopamine neurons of the rat A10 region: Molecular evidence for diverse signaling properties. *Brain Structure and Function*, 218(5):1159—1176.

Lindgren, N., Usiello, A., Goiny, M., Haycock, J., Erbs, E., Greengard, P., Hökfelt, T., Borrelli, E., and Fisone, G. (2003). Distinct roles of dopamine D2L and D2S receptor isoforms in the regulation of protein phosphorylation at presynaptic and postsynaptic sites. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7):4305–4309.

Lindvall, O. and Björklund, A. (1979). Dopaminergic innervation of the globus pallidus by collaterals from the nigrostriatal pathway. *Brain Research*, 172(1):169–173.

López-Cruz, L., Salamone, J. D., and Correa, M. (2018). Caffeine and selective adenosine receptor antagonists as new therapeutic tools for the motivational symptoms of depression. *Frontiers in Pharmacology*, 9(526):1–14.

Mackintosh, N. J. (1974). *The psychology of animal learning*. Academic Press, Oxford.

Mannella, F., Gurney, K., and Baldassarre, G. (2013). The nucleus accumbens as nexus between values and goals in goal-directed behavior: review and a new hypothesis. *Frontiers in Behavioral Neuroscience*, 7(135):1–29.

Mar, A. C., Horner, A. E., Nilsson, S. R. O., Alsiö, J., Kent, B. A., Kim, C. H., Holmes, A., Saksida, L. M., and Bussey, T. J. (2013). The touchscreen operant platform for assessing executive function in rats and mice. *Nature Protocols*, 8(10):1985–2005.

Marcott, P. F., Mamaligas, A. A., and Ford, C. P. (2014). Phasic dopamine release drives rapid activation of striatal D2-receptors. *Neuron*, 84(1):164–176.

Matamales, M., Bertran-Gonzalez, J., Salomon, L., Degos, B., Deniau, J. M., Valjent, E., Hervé, D., and Girault, J. A. (2009). Striatal medium-sized spiny neurons: Identification by nuclear staining and study of neuronal subpopulations in BAC transgenic mice. *PLoS ONE*, 4(3).

McAlonan, K. and Brown, V. J. (2003). Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behavioural Brain Research*, 146(1-2):97–103.

McFarland, B. R. and Klein, D. N. (2009). Emotional reactivity in depression: Diminished responsiveness to anticipated reward but not to anticipated punishment or to nonreward or avoidance. *Depression and Anxiety*, 26(2):117–122.

Mehta, M. A., Swainson, R., Ogilvie, A. D., Sahakian, B., and Robbins, T. W. (2001). Improved short-term spatial memory but impaired reversal learning following the dopamine D2 agonist bromocriptine in human volunteers. *Psychopharmacology*, 159(1):10–20.

Meredith, G. E., Agolia, R., Arts, M. P., Groenewegen, H. J., and Zahm, D. S. (1992). Morphological differences between projection neurons of the core and shell in the nucleus accumbens of the rat. *Neuroscience*, 50(1):149–162.

Millan, M. J., Maiofiss, L., Cussac, D., Audinot, V., Boutin, J. A., and Newman-Tancredi, A. (2002). Differential actions of antiparkinson agents at multiple classes of monoaminergic receptor. I. A multivariate analysis of the binding profiles of 14 drugs at 21 native and cloned human receptor subtypes. *Journal of Pharmacology and Experimental Therapeutics*, 303(2):791–804.

Millan, M. J., Newman-Tancredi, A., Quentric, Y., and Cussac, D. (2001). The "selective" dopamine D1 receptor antagonist, SCH23390, is a potent and high efficacy agonist at cloned human serotonin2C receptors. *Psychopharmacology*, 156(1):58–62.

Milton, L. K., Mirabella, P. N., Greaves, E., Spanswick, D. C., Van, M., Buuse, D., Oldfield, B. J., and Foldi, C. J. (2020). Suppression of cortico-striatal circuit activity improves cognitive flexibility and prevents body weight loss in activity-nbased anorexia in rats. *Biological Psychiatry*, Accepted.

Mingote, S., Amsellem, A., Kempf, A., Rayport, S., and Chuhma, N. (2019). Neurochemistry international dopamine-glutamate neuron projections to the nucleus accumbens medial shell and behavioral switching. *Neurochemistry International*, 129(104482):1–12.

Mingote, S., Chuhma, N., Kusnoor, S. l. V., Field, B., Deutch, A. Y., and Rayport, S. (2015). Functional connectome analysis of dopamine neuron glutamatergic connections in forebrain regions. *The Journal of Neuroscience*, 35(49):16259–16271.

Mishkin, M. and Pribram, K. H. (1955). Analysis of the effects of frontal lesions in monkeys: I. Variations of delayed alternations. *Journal of Comparative and Physiological Psychology*, 48:492–495.

Missale, C., Russel Nash, S., Robinson, S. W., Jaber, M., and Caron, M. G. (1998). Dopamine receptors: From structure to function. *Physiological Reviews*, 78(1):189–225.

Mogenson, G. J., Jones, D. L., and Yim, C. Y. (1980). From motivation to action: Functional interface between the limbic system and the motor system. *Progress in Neurobiology*, 14(2-3):69–97.

Mohebi, A., Pettibone, J. R., Hamid, A. A., Wong, J.-m. T., Vinson, L. T., Patriarchi, T., Tian, L., Kennedy, R. T., and Berke, J. D. (2019). Dissociable dopamine dynamics for learning and motivation. *Nature*, 570:65–70.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3):134–140.

Montagu, K. A. (1957). Catechol compounds in rat tissues and in brains of different animals. *Nature*, 180:244–245.

Morales, M. and Margolis, E. B. (2017). Ventral tegmental area: Cellular heterogeneity, connectivity and behaviour. *Nature Reviews Neuroscience*, 18(2):73–85.

Moreno, E., Chiarlone, A., Medrano, M., Puigdellívol, M., Bibic, L., Howell, L. A., Resel, E., Puente, N., Casarejos, M. J., Perucho, J., Botta, J., Suelves, N., Ciruela, F., Ginés, S., Galve-Roperh, I., Casadó, V., Grandes, P., Lutz, B., Monory, K., Canela, E. I., Lluís, C., McCormick, P. J., and Guzmán, M. (2018). Singular location and signaling profile of adenosine A2A-cannabinoid CB1 receptor heteromers in the dorsal striatum. *Neuropsychopharmacology*, 43(5):964–977.

Morris, M. C., Evans, L. D., Rao, U., and Garber, J. (2015). Executive function moderates the relation between coping and depressive symptoms. *Anxiety, Stress and coping*, 28(1):31–49.

Mrzljak, L., Bergson, C., Pappy, M., Huff, R., Levenson, R., and Goldman-Rakic, P. S. (1996). Localisation of dopamine D4 receptors in GABAergic neurons of the primate brain. *Nature*, 381:245–248.

Muller, A., Joseph, V., Slesinger, P. A., and Kleinfeld, D. (2014). Cell-based reporters reveal in vivo dynamics of dopamine and norepinephrine release in murine cortex. *Nature Methods*, 11(12):1245–1252.

Nichols, C. D. and Roth, B. L. (2009). Engineered G-protein coupled receptors are powerful tools to investigate biological processes and behaviors. *Frontiers in molecular neuroscience*, 2(10):16.

Nicola, S. M., Hopf, F. W., and Hjelmstad, G. O. (2004a). Contrast enhancement: A physiological effect of striatal dopamine? *Cell and Tissue Research*, 318:93–106.

Nicola, S. M., Yun, I. A., Wakabayashi, K. T., and Fields, H. L. (2004b). Cue-evoked firing of nucleus accumbens neurons encodes motivational significance during a discriminative stimulus task. *Journal of Neurophysiology*, 91(4):1840–1865.

Nilsson, S. R. O., Alsiö, J., Somerville, E. M., and Clifton, P. G. (2015). The rat's not for turning: Dissociating the psychological components of cognitive inflexibility. *Neuroscience and Biobehavioral Reviews*, 56(October 1980):1–14.

Niv, Y., Edlund, J. A., Dayan, P., and O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *The Journal of Neuroscience*, 32(2):551–562.

Nunes, E. J., Randall, P. A., Hart, E. E., Freeland, C., Yohn, S. E., Baqi, Y., Muller, C. E., Lopez-Cruz, L., Correa, M., and Salamone, J. D. (2013). Effort-related motivational effects of the VMAT-2 inhibitor tetrabenazine: Implications for animal models of the motivational symptoms of depression. *The Journal of Neuroscience*, 33(49):19120–19130.

O'Doherty, J. P. (2011). Chapter 14: Reward predictions and computations. In *Gottfried JA, ed. Neurobiology of Sensation and Reward. Boca Raton (FL): CRC Press/Taylor & Francis*. CRC Press.

O'Doherty, J. P., Buchanan, T. W., Seymour, B., and Dolan, R. J. (2006). Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron*, 49(1):157–166.

O'Doherty, J. P., Cockburn, J., and Pauli, W. M. (2017). Learning, reward, and decision making. *Annual Review of Psychology*, 68(1):73–100.

O'Neill, M. and Brown, V. J. (2007). The effect of striatal dopamine depletion and the adenosine A2A antagonist KW-6002 on reversal learning in rats. *Neurobiology of Learning and Memory*, 88(1):75–81.

Oomen, C. A., Hvoslef-Eide, M., Heath, C. J., Mar, A. C., Horner, A. E., Bussey, T. J., and Saksida, L. M. (2013). The touchscreen operant platform for testing working memory and pattern separation in rats and mice. *Nature protocols*, 8(10):2006–21.

Ostlund, S. B. and Balleine, B. W. (2008). On habits and addiction: an associative analysis of compulsive drug seeking. *Drug Discovery Today: Disease Models*, 5(4):235–245.

Pallanti, S., Hollander, E., and Goodman, W. K. (2004). A qualitative analysis of nonresponse: Management of treatment-refractory Obsessive-Compulsive Disorder. *Journal of Clinical Psychiatry*, 65:6–10.

Parker, J. G., Zweifel, L. S., Clark, J. J., Evans, S. B., Phillips, P. E., and Palmiter, R. D. (2010). Absence of NMDA receptors in dopamine neurons attenuates dopamine release but not conditioned approach during Pavlovian conditioning. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30):13491–13496.

Parkinson, J. A., Cardinal, R. N., and Everitt, B. J. (2000). Limbic cortical-ventral striatal systems underlying appetitive conditioning. *Progress in Brain Research*, 126:263–285.

Parsons, L. H. and Justice, J. B. (1992). Extracellular concentration and in vivo recovery of dopamine in the nucleus accumbens using microdialysis. *Journal of Neurochemistry*, 58(1):212–218.

Patriarchi, T., Cho, J. R., Merten, K., Howe, M. W., Marley, A., Xiong, W. H., Folk, R. W., Broussard, G. J., Liang, R., Jang, M. J., Zhong, H., Dombeck, D., von Zastrow, M., Nimmerjahn, A., Gradinaru, V., Williams, J. T., and Tian, L. (2018). Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science*, 360(6396):1–8.

Paxinos, G. and Watson, C. (1998). *The rat brain in stereotaxic coordinates*. Academic Press.

Peters, J., LaLumiere, R. T., and Kalivas, P. W. (2008). Infralimbic prefrontal cortex is responsible for inhibiting cocaine seeking in extinguished rats. *The Journal of Neuroscience*, 28(23):6046–6053.

Peterson, D. A., Elliott, C., Song, D. D., Makeig, S., Sejnowski, T. J., and Poizner, H. (2009). Probabilistic reversal learning is impaired in Parkinson's disease. *Neuroscience*, 163(4):1092–1101.

Phillips, P. E., Stuber, G. D., Helen, M. L., Wightman, R. M., and Carelli, R. M. (2003). Subsecond dopamine release promotes cocaine seeking. *Nature*, 422(6932):614–618.

Piantadosi, P. T., Yeates, D. C. M., and Floresco, S. B. (2018). Cooperative and dissociable involvement of the nucleus accumbens core and shell in the promotion and inhibition of actions during active and inhibitory avoidance. *Neuropharmacology*, 138:57–71.

Pisani, A., Bonsi, P., Centonze, D., Calabresi, P., and Bernardi, G. (2000). Activation of D2-like dopamine receptors reduces synaptic inputs to striatal cholinergic interneurons. *The Journal of Neuroscience*, 20(7):6–11.

Prediger, R. D. (2010). Effects of caffeine in Parkinson's disease: From neuroprotection to the management of motor and non-motor symptoms. *Journal of Alzheimer's Disease*, 20:205–220.

Qi, J., Zhang, S., Wang, H.-l., Barker, D. J., Miranda-barrientos, J., Morales, M., Section, N. N., Neuroscience, I., Abuse, D., and States, U. (2016). VTA glutamatergic inputs to nucleus accumbens drive aversion by acting on GABAergic interneurons. *Nature Methods*, 19(5):725–733.

Radke, A. K., Kocharian, A., Covey, D. P., Lovinger, D. M., Cheer, J. F., Mateo, Y., and Holmes, A. (2018). Contributions of nucleus accumbens dopamine to cognitive flexibility. *European Journal of Neuroscience*, 50(3):2013–2035.

Ragozzino, M. E. (2007). The contribution of the medial prefrontal cortex, orbitofrontal cortex, and dorsomedial striatum to behavioral flexibility. *Annals of the New York Academy of Sciences*, 1121:355–375.

Ragozzino, M. E. and Choi, D. (2004). Dynamic changes in acetylcholine output in the medial striatum during place reversal L learning. *Learning & Memory*, 11(1):70–77.

Ragozzino, M. E., Jih, J., and Tzavos, A. (2002). Involvement of the dorsomedial striatum in behavioral flexibility: Role of muscarinic cholinergic receptors. *Brain Research*, 953(1-2):205–214.

Reddy, L. F., Waltz, J. A., Green, M. F., Wynn, J. K., and Horan, W. P. (2016). Probabilistic reversal learning in schizophrenia: Stability of deficits and potential causal mechanisms. *Schizophrenia Bulletin*, 42(4):942–951.

Redfern, W. S., Tse, K., Grant, C., Keerie, A., Simpson, D. J., Pedersen, J. C., Rimmer, V., Leslie, L., Klein, S. K., Karp, N. A., Sillito, R., Chartsias, A., Lukins, T., Heward, J., Vickers, C., Chapman, K., and Armstrong, J. D. (2017). Automated recording of home cage activity and temperature of individual rats housed in social groups: The Rodent Big Brother project. *PLoS ONE*, 12(9):1–26.

Redgrave, P. and Gurney, K. (2006). The short-latency dopamine signal: A role in discovering novel actions? *Nature Reviews Neuroscience*, 7(12):967–975.

Remijnse, P. L., Nielen, M. M., Van Balkom, A. J., Cath, D. C., Van Oppen, P., Uylings, H. B., and Veltman, D. J. (2006). Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. *Archives of General Psychiatry*, 63:1225–1236.

Remijnse, P. L., van den Heuvel, O. A., Nielen, M. M. A., Vriend, C., Hendriks, G. J., Hoogendijk, W. J. G., Uylings, H. B. M., and Veltman, D. J. (2013). Cognitive onflexibility in obsessive-compulsive disorder and major depression is associated with distinct neural correlates. *PLoS ONE*, 8(4:e59600):1–9.

Rescorla, R. A. and Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement BT - Clasical conditioning II: current research and theory. *Clasical conditioning II: current research and theory*, pages 64–99.

Richfield, E. K., Penney, J. B., and Young, A. B. (1989). Anatomical and affinity state comparisons between dopamine D1 and D2 receptors in the rat central nervous system. *Neuroscience*, 30(3):767–777.

Richtand, N. M., Kelsoe, J. R., Segal, D., and Kuczenski, R. (1995). Regional quantification of D1, D2, and D3 dopamine receptor mRNA in rat brain using a ribonuclease protection assay. *Molecular Brain Research*, 33:97–103.

Rivera, A., Alberti, I., Martín, A. B., Narváez, J. A., De la Calle, A., and Moratalla, R. (2002). Molecular phenotype of rat striatal neurons expressing the dopamine D5 receptor subtype. *European Journal of Neuroscience*, 16(11):2049–2058.

Robbins, T. W. and Everitt, B. J. (1992). Functions of dopamine in the dorsal and ventral striatum. *Seminars in Neuroscience*, 4(2):119–127.

Robbins, T. W., Gillan, C. M., Smith, D. G., Wit, S. D., and Ersche, K. D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: Towards dimensional psychiatry. *Cell Press*, 16(1):81–91.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., and Rabbitt, P. (1994). Cambridge neuropsychological test automated battery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia*, 5(5):266–281.

Roberts, A. C., Tomic, D. L., Parkinson, C. H., Roeling, T. A., Cutter, D. J., Robbins, T. W., and Everitt, B. J. (2007). Forebrain connectivity of the prefrontal cortex in the marmoset monkey (Callithrix jacchus): An anterograde and retrograde tract-tracing study. *The Journal of Comparative Neurology*, 502(1):86–112.

Robinson, O. J., Cools, R., Carlisi, C. O., Sahakian, B. J., and Drevets, W. C. (2012). Ventral striatum responses during reward and punishment reversal learnign in unmedicated major deppressive disorder. *American Journal of Psychiatry*, 169:152–159.

Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., and Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Journal of Cognitive Neuroscience*, 12(1):142–162.

Romera-Garcia, R., Hook, R. W., Tiego, J., Bethlehem, R. A. I., Goodyer, I. M., Jones, P. B., Dolan, R., Grant, J. E., Bullmore, E. T., Yücel, M., and Chamberlain, S. R. (2020). Brain micro-architecture and disinhibition: A latent phenotyping study across 33 impulsive and compulsive behaviours. *Neuropsychopharmacology*, 0:1–9.

Rygula, R., Clarke, H. F., Cardinal, R. N., Cockcroft, G. J., Xia, J., Dalley, J. W., Robbins, T. W., and Roberts, A. C. (2015). Role of central serotonin in anticipation of rewarding and punishing outcomes: Effects of selective amygdala or orbitofrontal 5-HT depletion. *Cerebral Cortex*, 25(9):3064–3076.

Rygula, R., Noworyta-Sokolowska, K., Drozd, R., and Kozub, A. (2018). Using rodents to model abnormal sensitivity to feedback in depression. *Neuroscience and Biobehavioral Reviews*, 95(May):336–346.

Sahakian, B. J., Owen, A. M., Morant, N. J., Eagger, S. A., Boddington, S., Crayton, L., Crockford, H. A., Crooks, M., Hill, K., and Levy, R. (1993). Further analysis of the cognitive effects of tetrahydroaminoacridine (THA) in Alzheimer's disease: assessment of attentional and mnemonic function using CANTAB. *Psychopharmacology*, 110(4):395–401.

Sala-Bayo, J., Fiddian, L., Nilsson, S. R., Hervig, M. E., McKenzie, C., Mareschi, A., Boulos, M., Zhukovsky, P., Nicholson, J., Dalley, J. W., Alsiö, J., and Robbins, T. W. (2020). Dorsal and ventral striatal dopamine D1 and D2 receptors differentially modulate distinct phases of serial visual reversal learning. *Neuropsychopharmacology*, 45:736 – 744.

Salamone, J. D., Correa, M., Farrar, A. M., Nunes, E. J., and Pardo, M. (2009). Dopamine, behavioral economics and effort. *Frontiers in Behavioral Neuroscience*, 3(13):1–12.

Salamone, J. D., Correa, M., Ferrigno, S., Yang, J.-H., Rotolo, R. A., and Presby, R. E. (2018). The psychopharmacology of effort-related decision making: Dopamine, adenosine, and insights into the neurochemistry of motivation. *Pharmacological Reviews*, 70(10):747–762.

Salery, M., Trifilieff, P., Caboche, J., and Vanhoutte, P. (2020). From signaling molecules to circuits and behaviors: Cell-type–specific adaptations to psychostimulant exposure in the striatum. *Biological Psychiatry*, 87(11):944–953.

Salgado, S. and Kaplitt, M. G. (2015). The nucleus accumbens: A comprehensive review. *Stereotactic and Functional Neurosurgery*, 93(2):75–93.

Santerre, J. L., Nunes, E. J., Kovner, R., Leser, C. E., Randall, P. A., Collins-Praino, L. E., Lopez Cruz, L., Correa, M., Baqi, Y., Müller, C. E., and Salamone, J. D. (2012). The novel adenosine A2A antagonist prodrug MSX-4 is effective in animal models related to motivational and motor functions. *Pharmacology Biochemistry and Behavior*, 102(4):477–487.

Sato, Y., Endo, H., Ajiki, T., Hakamata, Y., Okada, T., Murakami, T., and Kobayashi, E. (2004). Establishment of Cre/LoxP recombination system in transgenic rats. *Biochemical and Biophysical Research Communications*, 319(4):1197–1202.

Saunders, B. T., Richard, J. M., Margolis, E. B., and Janak, P. H. (2018). Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nature Neuroscience*, 21(8):1072–1083.

Sautel, F., Griffon, N., Sokoloff, P., Schwartz, J.-C., Launay, C., Simon, P., Costentin, J., Schoenfelderr, A., Garrido, F., Mann, A., and Wermuth, C. G. (1995). Nafadotride, a potent preferential dopamine D3 receptor antagonist, activates locomotion in Rodents. *The Journal of Pharmacology and Experimental Therapeutics*, 275(3):1239–1246.

Savasta, M., Dubois, A., and Scatton, B. (1986). Autoradiographic localization of D1 dopamine receptors in the rat brain with [3H]SCH 23390. *Brain Research*, 375(2):291–301.

Schiffmann, S. N., Fisone, G., Moresco, R., Cunha, R. A., and Ferré, S. (2007). Adenosine A2A receptors and basal ganglia physiology. *Progress in Neurobiology*, 83:277–292.

Schoenbaum, G. and Setlow, B. (2003). Lesions of nucleus accumbens disrupt learning about aversive outcomes. *The Journal of Neuroscience*, 23(30):9833–9841.

Schultz, W. (2013). Updating dopamine reward signals. *Current Opinion in Neurobiology*, 23:229–238.

Schultz, W. (2019). Recent advances in understanding the role of phasic dopamine activity. *F1000Research*, 8:1680.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599.

Schwartz, J.-C., Diaz, J., Bordet, R., Griffon, N., Perachon, S., Pilo, C., Ridray, S., and Sokoloff, P. (1998). Functional implications of multiple dopmaine receptor subtypes: the D1/D3 receptor coexistence. *Brain research reviews*, 26:236–242.

Schwartz, J.-C., Diaz, J., Griffon, N., Lammers, C., Lévesque, D., Martres, M. P., and Sokoloff, P. (1994). The dopamine D3 receptor in nucleus accumbens: selective cellular localisation, function and regulation. *European Neuropsychopharmacology*, 4(3):190–191.

Serpell, L., Waller, G., Fearon, P., and Meyer, C. (2009). The roles of persistence and perseveration in psychopathology. *Behavior Therapy*, 40:260–271.

Sesack, S. R., Deutch, A. Y., Roth, R. H., and Bunney, B. S. (1989). Topographical organization of the efferent projections of the medial prefrontal cortex in the rat: An anterograde tract-tracing study with Phaseolus vulgaris Leucoagglitinin. *The Joirnal of Comparative Neurology*, 290:213–242.

Sesia, T., Temel, Y., Lim, L. W., Blokland, A., Steinbusch, H. W., and Visser-Vandewalle, V. (2008). Deep brain stimulation of the nucleus accumbens core and shell: opposite effects on impulsive action. *Experimental Neurology*, 214(1):135–139.

Shah, D., Verhoye, M., Van der Linden, A., and D'Hooge, R. (2019). Acquisition of spatial search strategies and reversal learning in the Morris water maze depend on disparate brain functional connectivity in mice. *Cerebral Cortex*, 29:4519–4529.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century-Crofts.

Smith, A. G., Neill, J. C., and Costall, B. (1999). The dopamine D3/D2 receptor agonist 7-OH-DPAT induces cognitive impairment in the marmoset. *Pharmacology Biochemistry and Behavior*, 63(2):201–211.

Smith, I. D. and Grace, A. A. (1992). Role of the subthalamic nucleus in the regulation of nigral dopamine neuron activity. *Synapse*, 12(4):287–303.

Soares-Cunha, C., Coimbra, B., Sousa, N., and Rodrigues, A. J. (2016). Reappraising striatal D1- and D2- neurons in reward and aversion. *Neuroscience and Biobehavioral Reviews*, 68:370–386.

Sohn, M. H., Ursu, S., Anderson, J. R., Stenger, V. A., and Carter, C. S. (2000). The role of prefrontal cortex and posterior parietal cortex in task switching. *Proceedings of the National Academy of Sciences of the United States of America*, 97(24):13448–13453.

Somogyi, P., Bolam, J. P., and Smith, A. D. (1981). Monosynaptic cortical input and local axon collaterals of identified striatonigral neurons. A light and electron microscopic study using the golgi-peroxidase transport-degeneration procedure. *Journal of Comparative Neurology*, 195(4):567–584.

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., and Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, 16(7):966–973.

Steiner, H. and Gerfen, C. R. (1998). Role of dynorphin and enkephalin in the regulation of striatal output pathways and behavior. *Experimental Brain Research*, 123(1-2):60–76.

Stern, C. E. and Passingham, R. E. (1995). The nucleus accumbens in monkeys (Macaca fascicularis). III. Reversal learning. *Experimental Brain Research*, 106(2):239–247.

Stuber, G. D., Hnasko, T. S., Britt, J. P., Edwards, R. H., and Bonci, A. (2010). Dopaminergic terminals in the nucleus accumbens but not the dorsal striatum corelease glutamate. *Journal of Neuroscience*, 30(24):8229–8233.

Surguladze, S., Brammer, M. J., Keedwell, P., Giampietro, V., Young, A. W., Travis, M. J., Williams, S. C., and Phillips, M. L. (2005). A differential pattern of neural response toward sad versus happy facial expressions in major depressive disorder. *Biological Psychiatry*, 57(3):201–209.

Surmeier, D. J., Ding, J., Day, M., Wang, Z., and Shen, W. (2007). D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends in Neurosciences*, 30(5):228–235.

Surmeier, D. J., Eberwine, J., Wilson, C. J., Cao, Y., Stefani, A., and Kitai, S. T. (1992). Dopamine receptor subtypes colocalize in rat striatonigral neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 89:10178–10182.

Surmeier, D. J., Plotkin, J., and Shen, W. (2009). Dopamine and synaptic plasticity in dorsal striatal circuits controlling action selection. *Current Opinion in Neurobiology*, 19(6):621–628.

Surmeier, D. J., Song, W. J., and Yan, Z. (1996). Coordinated expression of dopamine receptors in neostriatal medium spiny neurons. *The Journal of Neuroscience*, 16(20):6579–6591.

Swainson, R., Rogers, R. D., Sahakian, B. J., Summers, B. A., Polkey, C. E., and Robbins, T. W. (2000). Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: Possible adverse effects of dopaminergic medication. *Neuropsychologia*, 38(5):596–612.

Swanson, C. J., Heath, S., Stratford, T. R., and Kelley, A. E. (1997). Differential behavioral responses to dopaminergic stimulation of nucleus accumbens subregions in the rat. *Pharmacology Biochemistry and Behavior*, 58(4):933–945.

Syed, E. C. J., Grima, L. L., Magill, P. J., Bogacz, R., Brown, P., and Walton, M. E. (2015). Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nature Neuroscience*, 19(1):34–36.

Szechtman, H., Sulis, W., and Eilam, D. (1998). Quinpirole induces compulsive checking behavior in rats: A potential animal model of obsessive-compulsive disorder (OCD). *Behavioural Neuroscience*, 112(6):1475–1485.

Taghzouti, K., Le Moal, M., and Simon, H. (1985). Enhanced frustrative nonreward effect following 6-hydroxydopamine lesions of the lateral septum in the rat. *Behavioral Neuroscience*, 99(6):1066–1073.

Takahashi, Y. K., Langdon, A. J., Niv, Y., and Schoenbaum, G. (2016). Temporal specificity of reward prediction errors signaled by putative dopamine neurons in rat VTA depends on ventral striatum. *Neuron*, 91(1):182–193.

Takahashi, Y. K., Roesch, M. R., Stalnaker, T. A., Haney, R. Z., Calu, D. J., Taylor, A. R., Burke, K. A., and Schoenbaum, G. (2009). The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron*, 62(2):269–280.

Taylor, J. R. and Robbins, T. W. (1984). Enhanced behavioural control by conditioned reinforcers following microinjections of d-amphetamine into the nucleus accumbens. *Psychopharmacology*, 84(3):405–412.

Taylor, J. R. and Robbins, T. W. (1986). 6-Hydroxydopamine lesions of the nucleus accumbens, but not of the caudate nucleus, attenuate enhanced responding with reward-related stimuli produced by intra-accumbens d-amphetamine. *Psychopharmacology*, 90(3):390–397.

Tecuapetla, F., Patel, J. C., Xenias, H., English, D., Tadros, I., Shah, F., Berlin, J., Deisseroth, K., Rice, M. E., Tepper, J. M., and Koos, T. (2010). Glutamatergic signaling by mesolimbic dopamine neurons in the nucleus accumbens. *The Journal of Neuroscience*, 30(20):7105–7110.

Tepper, J. M. and Bolam, J. P. (2004). Functional diversity and specificity of neostriatal interneurons. *Current Opinion in Neurobiology*, 14(6):685–692.

Tobler, P. N., Dickinson, A., and Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *The Journal of Neuroscience*, 23(32):10402–10410.

Tremblay, L. K., Naranjo, C. A., Cardenas, L., Herrmann, N., and Busto, U. E. (2002). Probing brain reward system function in major depressive disorder. *Archives of General Psychiatry*, 59(5):409.

Tritsch, N. X. and Sabatini, B. L. (2012). Dopaminergic modulation of synaptic transmission in cortex and striatum. *Neuron*, 76(1):33–50.

Tunstall, M. J., Oorschot, D. E., Kean, A., and Wickens, J. R. (2002). Inhibitory interactions between spiny projection neurons in the rat striatum. *Journal of Neurophysiology*, 88(3):1263–1269.

Turner, K. M. and Parkes, S. L. (2020). Prefrontal regulation of behavioural control: Evidence from learning theory and translational approaches in rodents. *Neuroscience & Biobehavioral Reviews*, 118(2):27–41.

Usiello, A., Baik, J. H., Rougé-Pont, F., Picetti, R., Dierich, A., LeMeur, M., Piazza, P. V., and Borrelli, E. (2000). Distinct functions of the two isoforms of dopamine D2 receptors. *Nature*, 408:199–203.

Verharen, J. P., den Ouden, H. E., Adan, R. A., and Vanderschuren, L. J. (2020). Modulation of value-based decision making behavior by subregions of the rat prefrontal cortex. *Psychopharmacology*, 237(5):1267–1280.

Verharen, J. P. H. (2018). *Neuroeconomic measures of reward and aversion*. PhD thesis, Utrecht University.

Verharen, J. P. H., Adan, R. A. H., and Vanderschuren, L. J. M. J. (2019). Differential contributions of striatal dopamine D1 and D2 receptors to component processes of value-based decision making. *Neuropsychopharmacology*, 0:1–10.

Verharen, J. P. H., De Jong, J. W., Roelofs, T. J. M., Huffels, C. F. M., Van Zessen, R., Luijendijk, M. C. M., Hamelink, R., Willuhn, I., Den Ouden, H. E. M., Van Der Plasse, G., Adan, R. A., and Vanderschuren, L. J. M. J. (2018). A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. *Nature Communications*, 9(1):1–15.

Vertes, R. P. (2004). Differential Projections of the Infralimbic and Prelimbic Cortex in the Rat. *Synapse*, 51(1):32–58.

Volkow, N. D., Fowler, J. S., Wang, G. J., Baler, R., and Telang, F. (2009). Imaging dopamine's role in drug abuse and addiction. *Neuropharmacology*, 56:3–8.

Vontell, R., Segovia, K. N., Betz, A. J., Mingote, S., Goldring, K., Cartun, R. W., and Salamone, J. D. (2010). Immunocytochemistry studies of basal ganglia adenosine A2A receptors in rat and human tissue. *The Journal of Histotechnology*, 33(1):41–47.

Voorn, P., Vanderschuren, L. J., Groenewegen, H. J., Robbins, T. W., and Pennartz, C. M. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences*, 27(8):468–474.

Waelti, P., Dickinson, A., and Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, 412:43–48.

Walker, S. C., Robbins, T. W., and Roberts, A. C. (2009). Differential contributions of dopamine and serotonin to orbitofrontal cortex function in the marmoset. *Cerebral Cortex*, 19(4):889–898.

Wall, N. R., DeLaParra, M., Callaway, E. M., and Kreitzer, A. C. (2013). Differential innervation of direct- and indirect-pathway striatal projection neurons. *Neuron*, 79(2):347–360.

Waltz, J. A. and Gold, J. M. (2016). Motivational deficits in schizophrenia and the representation of expected value. *Current Topics in Behavioural Neuroscience*, 27:375–410.

Wanat, M. J., Willuhn, I., Clark, J. J., and Phillips, P. E. M. (2009). Phasic dopamien release in appetitive behviors and drug abuse. *Current Drug Abuse Review*, 2(2):195–213.

Wang, Z., Kai, L., Day, M., Ronesi, J., Yin, H. H., Ding, J., Tkatch, T., Lovinger, D. M., and Surmeier, D. J. (2006). Dopaminergic control of corticostriatal long-term synaptic depression in medium spiny neurons is mediated by cholinergic interneurons. *Neuron*, 50(3):443–452.

Wassum, K. M., Ostlund, S. B., and Maidment, N. T. (2012). Phasic mesolimbic dopamine signaling precedes and predicts performance of a self-initiated action sequence task. *Biological Psychiatry*, 71(10):846–854.

Weiner, I., Gal, G., Rawlins, J. N., and Feldon, J. (1996). Differential involvement of the shell and core subterritories of the nucleus accumbens in latent inhibition and amphetamine-induced activity. *Behavioural Brain Research*, 81:123–133.

Weiner, R. I. and Ganong, W. F. (1978). Role of brain monoamines and histamine in regulation of anterior pituitary secretion. *Physiological Reviews*, 58(4):905–976.

Wiecki, T. V., Sofer, I., and Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, 7(14):1–10.

Wilson, C. J. and Groves, P. M. (1980). Fine structure and synaptic connections of the common spiny neuron of the rat neostriatum: A study employing intracellular injection of horseradish peroxidase. *Journal of Comparative Neurology*, 194(3):599–615.

Wyvell, C. L. and Berridge, K. C. (2000). Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: Enhancement of reward 'wanting' without enhanced 'liking' or response reinforcement. *The Journal of Neuroscience*, 20(21):8122–8130.

Yapo, C., Nair, A. G., Clement, L., Castro, L. R., Kotaleski, J. H., and Vincent, P. (2017). Detection of phasic dopamine by D1 and D2 striatal medium spiny neurons. *The Journal of Physiology*, 24:7451–7475.

Yerkes, R. M. and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5):459–480.

Yin, H. H. and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature reviews. Neuroscience*, 7(6):464–476.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, 19:181–189.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., and Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience*, 22(2):513–523.

Young, L. (2019). *Neurochemical and neuroanatomical basis of reversal learning in the rat*. PhD thesis, University of Cambridge.

Zahm, D. S. (1999). Functional-anatomical implications of the nucleus accumbens core and shell subterritories. *Annals of the New York Academy of Sciences*, 877:113–128.

Zell, V., Steinkellner, T., Hollon, N. G., Jin, X., Zweifel, L. S., Hnasko, T. S., Zell, V., Steinkellner, T., Hollon, N. G., Warlow, S. M., Souter, E., Faget, L., Hunker, A. C., Jin, X., Zweifel, L. S., and Hnasko, T. S. (2020). VTA glutamate neuron activity drives positive reinforcement absent dopamine co-release. *Neuron*, 107:1–10.

Zhou, J., Wu, B., Lin, X., Dai, Y., Li, T., Zheng, W., Guo, W., Chen, X., and Chen, J.-F. (2019). Accumbal adenosine A2A receptors enhance cognitive flexibility by facilitating strategy shifting. *Frontiers in Cellular Neuroscience*, 13(130):1–14.

Zhou, Q.-Y., Lu, C., Olah, M. E., Johnson, R. A., Stiles, G. L., and Civelli, O. (1992). Molecular cloning and characterization of the human A3 adenosine receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 89:7432–7436.

Zhukovsky, P., Puaud, M., Jupp, B., Sala-Bayo, J., Alsiö, J., Xia, J., Searle, L., Morris, Z., Sabir, A., Giuliano, C., Everitt, B. J., Belin, D., Robbins, T. W., and Dalley, J. W. (2019). Withdrawal from escalated cocaine self-administration impairs reversal learning by disrupting the effects of negative feedback on reward exploitation: a behavioral and computational analysis. *Neuropsychopharmacology*, 44(13):2163–2173.