Neuronal circuits underlying learning of competing action strategies

A tale of many players

Ana Mafalda dos Reis de Abreu Rodrigues Vicente

Dissertation presented to obtain the Ph.D degree in Biology | Neuroscience

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
March, 2016

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

Neuronal circuits underlying learning of competing action strategies

A tale of many players

Ana Mafalda dos Reis de Abreu Rodrigues Vicente

Dissertation presented to obtain the Ph.D degree in Biology | Neuroscience

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by:

Oeiras, March, 2016

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

"I have not yet lost a feeling of wonder, and of delight, that this delicate motion should reside in all things around us, revealing itself only to him who looks for it. I remember, in the winter of our first experiments, just seven years ago, looking on snow with new eyes. There the snow lay around my doorstep – great heaps of protons quietly precessing in the earth's magnetic field. To see the world for a moment as something rich and strange is the private reward of many a discovery."

Edward Mills Purcell

The work presented in this thesis was supported by Fundação Champalimaud, Instituto Gulbenkian Ciencia, Fundação para a Ciencia e Tecnologia (grant SFRH / BD / 33942 / 2009), European Research Council (STG 243393 and COG 617142) and European Research Area Network (Neuron-II/001/2013 and Neuron-II/002/2013). The work was conducted under the supervision of Dr. Rui M. Costa, and the thesis committee supervision of Dr. Marta Moita and Dr. Susana Lima

TABLE OF CONTENTS

ACKNOWLEDGMENTS

"Keep your head up, keep your heart strong"
Ben Howard

This thesis marks the end of a very long and eventful journey, which would not have happened without the help and support of many people.

First and foremost I would like to thank my supervisor, Rui Costa, for so many things that it is hard to put it on paper. For his sharp insights, for his integrity, for his support, for his genuine care, both in work and in life. For being the scientist and the person I aspire to be one day. For making this project happen - it would have been impossible without him. For his direction of the lab - he gives us all the freedom to try, and the guidance to believe. Because, more than just a supervisor, he became a friend. And because when he told me that he saw me as a colleague, I truly believed that I could one day be a scientist, and not just a student. Thank you Rui, it has been an amazing journey!

To Grant. For being my home. For all of the amazing things we've done together, and all the others still to come. For single headedly saving our relationship by created a way of locking the bed sheets in place at night. Thanks (and sorry) for all the late nights and the early mornings, and for keeping me sane through this journey. You make me smile when I am too serious, and you make me laugh even harder when I am happy. You bring

out the best of me, and I'm glad I get to share the rest of this crazy life with you. To you I say: "it's a magical world (…), let's go exploring", together.

I also want to thank my loving, caring and amazing family. You are the reason I am here. Thank you so much to my grandmas and my grandpa for all the love you have always given me, I could not have asked for better (and I don't think better is actually possible). And I want to thank my parents. For everything! I am the person I am nowadays because of you, and I will never be able to thank you enough. You have always loved, believed in and trusted me, no matter the time or place. If I had to create my own parents, I could not have done such a good job. The reason I can fly today is because you have always made sure I was never afraid of falling (and, just in case, you have always been there to catch me). I know how proud you are of me, but I sincerely hope you realize how proud I am of you, and of calling you my parents.

A big, huge thanks to my all friends for their support and encouragement, but in particular Ana P., Ana V., Anna, Andreia, Joana O., João, Kiersten, Libbi, Nicco, Pedro, Raquel, Susana, Vania and Vicky. This adventure has been a long and mutual one, and it has been a pleasure to share it all - highs and lows, smiles and tears, coffee and wine, days and nights - with you. Lets keep on going together, I hear the best times are still ahead of us.

I would also like to thank everyone in the Champalimaud Neuroscience Programme. I doubt that I will ever find a place with a similar spirit. "Work hard, play hard" will never be the same after working here, with this amazing group of people. Curiosity and the sense of wonder bind us all, and I sincerely hope that will never change. A special thank you goes to

my INDP year – we shared tons together, and the beer hunter will never be the same without us; and the Neurobiology of Action Lab – past and present members, and Pedro and Ana Vaz in particular, my long lasting partners in crime – for all the help, but also for the long discussions about almost everything, from basal ganglia functioning to how a zombie-inducing disease would be biologically possible. Thank you to my thesis committee, Marta Moita and Susana Lima, for all the help and support along this ride.

To finish, I'd like to acknowledge what I can only call sheer, damn luck. I am so thankful that I was born in a time and place where I was allowed to study, and to thrive, and where my education, my thoughts and my views matter. I am lucky that I've always had a roof over my head and food on my plate, and that my support system was always loving and secure. I am lucky that my home was never damaged by bombs, or destroyed by rotten ideologies. We are all in some way, perhaps more than we'd like to admit, products of luck. So, this last paragraph is more a wish than anything else. I hope that into the future dumb luck plays less and less of a role, and everyone lives a life with what I'm here calling luck, but should really just be called the bare minimum.

RESUMO

O sistema nervoso é uma adaptação incrível que permite aos animais reagir (e agir) de acordo com a sua percepção do ambiente que os rodeia. Portanto, uma característica extremamente importante de um organismo é a capacidade de aprender novas acções quando exposto a novos contextos, situações ou necessidades. Quando se aprende uma nova acção, o comportamento é altamente dependente do valor do seu resultado. No entanto, com múltiplas repetições, uma acção intencional pode tornar-se menos dependente do valor das suas consequências, e progredir para um hábito. Os hábitos são, geralmente, caracterizados pela relativa independência das suas consequências, e vistos como associações entre estímulos e respostas. Vários estudos apontam para que estes dois modos de comportamento dependem de circuitos dissociáveis no cérebro, com a zona medial do estriado dorsal e a amígdala baso-lateral envolvidos em acções intencionais, e o estriado dorso-lateral envolvido no controlo de hábitos. No entanto, pouco se sabe sobre a sua interacção durante a aprendizagem, ou como o controlo do comportamento é equilibrado pelo sistema. Adicionalmente, a dopamina, que é um importante neuromodulador, foi também implicada neste processo. Lesões do sistema dopaminérgico no estriado dorso-lateral, por exemplo, resultam em animais que aprendem acções de forma mais intencional.

Na primeira parte deste estudo, explorámos o papel do transportador de dopamina - responsável pela sua reabsorção - no equilíbrio entre as acções intencionais e habituais. Descobrimos que murganhos com níveis inferiores de expressão deste transportador formam hábitos num paradigma em que os controlos continuam sensíveis a manipulações no valor da recompensa da acção. Foi ainda verificado que a diferença não

advém da incapacidade de discriminação contextual, nem da impossibilidade de aprender acções intencionais. O transportador de dopamina parece, desta forma, ser responsável pela modulação da transição entre os dois modos de comportamento.

Na segunda parte deste trabalho foi investigada a importância das projecções da amígdala baso-lateral para o estriado. Treinando animais para pressionar uma alavanca para receber estimulação dos neurônios que projectam da amígdala para o estriado, observámos não só que os animais são capazes de aprender uma nova acção, mas também que a acção aprendida é sensível à degradação da contingência entre a mesma e o seu resultado. Descobrimos também que, ao contrário do que foi previamente observado, a amígdala baso-lateral projecta para as duas principais populações neuronais do estriado dorsal.

Por último, decidimos estudar essas duas populações de neurónios do estriado. O estriado é o principal núcleo de entrada de informação nos gânglios da base, recebendo aferentes glutamatérgicos de estruturas corticais, talamicas e amigdalares; a parte interna do *globus pallidus* e a *substantia nigra pars reticulata*, por outro lado, são os principais centros de saída de informação. Existem duas principais vias que ligam estas estruturas: os neurônios da via direta, que expressam o receptor D1 e projectam diretamente para a região de saída; e os neurónios da via indirecta, que expressam o receptor D2 e transmitem informação através de uma projecção multi-sináptica para a região de saída. Nós mostramos que, ao contrário do que seria classicamente previsto, a activação de ambas as vias é suficiente para reforçar uma acção, mas cada via suporta diferentes estratégias de acção.

Estas observações contribuem para a forma como se pensa no funcionamento dos gânglios da base durante o controlo de acções, e podem ter implicações importantes para doenças de natureza compulsiva.

SUMMARY

The nervous system is an amazing adaptation that allows animals to react to (and act on) their perception of the environment. One extremely important feature of an organism is the capacity to learn new actions upon exposure to new contexts, situations or necessities. When an action begins to be learned, the behaviour is highly dependent on the expected value of the outcome associated with it. However, with repetition, the execution of an action can become less dependent on the value of its consequences, and progress to a habit. Habitual behaviour is, in general, characterized by the relative independence from the outcome, and seen instead as a stimulus-response association. Several studies have pinpointed that these two behavioural modes depend on dissociable circuits in the brain, with dorsomedial striatum and basolateral amygdala involved in goal-directed actions, and dorsolateral striatum involved in habits. Nevertheless, little is known about their interaction during action learning, or how the system balances both modes during the behaviour output. Furthermore, dopamine, which is an important neuromodulator, has also been associated with this process. Accordingly, dopamine depletion in dorsolateral striatum leads to increased sensitivity to changes in the value of the outcome.

In the first part of these series of studies, we explored the role of the dopamine transporter – responsible for dopamine re-uptake – in the balance between goal-directed and habitual actions. We found that heterozygous knock out mice for that transporter became habitual in a training paradigm where littermate controls were goal-directed. We further discovered that such behaviour did not stem from a lack of context discrimination, nor inability to form goal-directed associations. Rather, it

appears the transporter is responsible for gating the switch between both behavioural modes.

In the second part of this work, we investigated the importance of the projections from basolateral amygdala to striatum. Since basolateral amygdala can track the value of stimuli associated with a reinforcer and is essential for goal-directed actions, we hypothesized the reinforcement signal needed to learn a new action might originate from this area. We trained animals to press a lever to receive stimulation of the neurons projecting from amygdala to dorsomedial striatum. Trained animals were able to learn the action, and their behaviour was highly sensitive to degradation of the contingency between action and outcome. We've also uncovered that, unlike previously expected, the basolateral amygdala is connected with the two main populations of projection neurons of striatum.

Finally, we set out to study the role of those striatal projection populations in goal-directed and habitual behaviour. The striatum is the main input nucleus of the basal ganglia, receiving glutamatergic inputs from cortical, thalamic and amygdalar structures, while the internal globus pallidus and the substantia nigra pars reticulata are the main output centres. There are two main pathways that connect these structures: the direct pathway neurons, which express D1 receptor and project directly to the output region; and the indirect pathway neurons, which express D2 receptor and relay the information through a multi-synaptic connection to the output region. It has been long hypothesised that these parallel projections have a dichotomous effect on movement and reinforcement of actions, with direct neurons signalling a go signal to appetitive stimuli, and indirect neurons a no-go response to aversive stimuli. We showed that self-stimulation of both

neuronal populations positively reinforced actions, but each pathway supported different action strategies.

Taken together, these results have important implications for understanding the role of basal ganglia circuits in controlling the balance for goal-directed and habitual actions, and have implications for understanding compulsive disorders.

ABBREVIATION LIST

AAV	Adeno-associated virus
AL	Active lever
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
AMPH	Amphetamine
BLA	Basolateral amygdala
CD	Contingency degradation
CeA	Central amygdala
ChR	ChannelRhodopsin
COMT	Catechol-O-methyltransferase
CPP	Conditioned place preference
CRF	Continuous reinforcement
CS	Conditional stimulus
DA	Dopamine
DAT	Dopamine transporter
$DAT^{+/-}$	Dopamine transporter knock-out heterozygous mice
$DAT^{+/+}$	Wild-type littermate mice
DLS	Dorsolateral striatum
DMS	Dorsomedial striatum
dMSNs	Direct pathway medium spiny neurons
EPSCs	Excitatory postsynaptic potential
FR	Fixed ratio
GFP	Green fluorescent protein
GPe	External globus pallidus
GPi	Internal globus pallidus
ICs	Intercalate cell nucleus
IL	Inactive lever

iMSNs	Indirect pathway medium spiny neurons
KO	Knock-out
LTD	Long-term depression
LTP	Long-term potentiation
MSNs	Medium spiny neurons
NAc	Nucleus accumbens
NMDAR	N-methyl-D-aspartate (NMDA) receptor
OCD	Obsessive compulsive disorder
OFC	Orbitofrontal cortex
PD	Parkinson's disease
RI	Random interval
RR	Random ratio
SNc	Substantia nigra pars compacta
SNr	Substantia nigra pars reticulata
STN	Subthalamic nucleus
US	Unconditional stimulus
VI	Variable interval
VR	Variable ratio
VTA	Ventral tegmental area
WT	Wild-type
YFP	Yellow fluorescent protein

1

INTRODUCTION

"Actions are inherently rational in a way that response can never be"

Dickinson, 1985

INTRODUCTION

Halfway through my PhD, our laboratory changed location. From then on, every time I wanted to reach the lab I would have to get into the building, go to the elevator, swipe my ID card and press the 2nd floor button. While in the beginning I would regularly press without swiping, which would lead nowhere, very rapidly I learned to systematically swipe before pressing. Thus, I could get there focusing on what I would have to do once I got there, rather than the actions that led me there. On the other hand, to leave the 2nd floor, an ID card is not necessary; one can simply push the lobby floor button. However, and once I knew about this, it took many attempts for me to finally stop swiping my card on the way out. Although it is clearly an unnecessary action for my desired outcome, I kept repeating it again and again for a longer time than I care to admit.

This story illustrates some of the questions that caught my attention from very early on in my Ph.D. It is very important in life to learn new things. Like me, early on going up in the elevator, we begin learning by trial and error, in a goal-directed manner. And habitual actions can be learned with the repetition of specific actions in similar contexts, such as always taking the elevator up when I get to the building. However, it is also very important to be able to not only switch back and forward between goal-directed actions and habits depending on your context, but also to go back to a goal-directed action if the habit is now non-adaptive (for example, unnecessarily ID swiping in the elevator). In this thesis we investigate the circuits that allow animals to not only learn a new, goal-directed action through a reinforcing signal, but also which mechanisms are in place to

provide a "normal" competition between circuitry involved in the control and balance of goal-directed and habitual actions.

A brief saga of goal-directed and habitual actions

Action, and learning new actions in particular, has always been a hot topic in philosophy, psychology and more recently neuroscience. Habit is an everyday word used intuitively to describe those behaviours that people do without "giving it too much thought". Several of the actions performed during the lifetime of an individual, or even in single day, can fit into this broad consideration of a habit. In a more precise view, a habit can be considered an action that became independent of the outcome, of its consequences, and is a response to a stimulus or situation. This contrast to those actions that we select based on their relationship to the outcome, which can be considered goal-directed.

Descartes, and his views on animal learning, strongly influenced the way we viewed animal behaviour in previous eras. Descartes defended that animals' actions can be simply described as a response to stimuli, purely by cause and effect laws, and instincts (Descartes and Clark, 1999; Descartes et al., 1985). He suggested that stimuli caused vibrations that could be sensed by sensor organs. Those sensor organs would lead to pulling of small "threads" which culminate with the release of fluid leading to the movement of muscle, and thus a reaction (Descartes et al., 1985).

According to Descartes, this could explain all non-human animals behaviour, as all non-voluntary human actions. In his view, the voluntary actions (only accessible to human beings) were controlled by the soul,

which would take over control of the movements of the body (Descartes and Clark, 1999; Descartes et al., 1985).

Although much changed since Descartes, the concept of action and animal learning as a series of stimulus and response associations remained for most of the 20th century. For many years during the last century the general concept of how animals learn an action was controlled by Hull's vision, based on the Thorndike's "Law of effect" - which states that events leading to a positive effect will be more likely to occur again, while responses that lead to negative effects will be less likely (Thorndike, 1898). According to Hull and colleges, an action was dependent upon a stimulus that is associated with a reward (Hull, 1943). Learning would thus depend on learning stimulus-response associations and on previous stimuli, and a new action could be viewed as a learned reflex. As a consequence, the action is seen as independent of the value of the outcome and not dependent of any representation of the goal in the brain.

A major opponent of this vision of learning and performing actions was Tolman, who introduced the concept of mental maps (Tolman, 1948): the brain has a representation of the environment (which is learned), and that representation allows for decisions that are not purely based on a response to a stimulus. Tolman and his colleagues designed an experiment to probe the existence of cognitive maps on rats, where they trained animals in a particular path to a reward, which was substituted by several radial arms on the testing day. They observed that the animals chose to walk through the arms that were closer to the reward site. These experiment demonstrated the presence of a broader mental map, not dependent solemnly on stimulus outcome learning, but on knowledge on food location.

Despite Tolman's ideas and experiments, it wasn't until late in the 20th century that experiments by Adams, Dickinson and Rescorla came to shine the light on this dichotomous view of learning. One of the first challenges was testing whether animals are simply capable of creating associations between a certain stimulus and a response (which are strengthen by a positive outcome), or whether a representation of the outcome is encoded in the associative knowledge of the behaviour output itself. Both theories are similar in which the action performed by the animals is not necessarily different. Therefore, what animals do does not necessarily reflect why the animals are doing it. It is necessary to test the consequentiality of the action. And in order to probe the associative structure being formed, one has to manipulate it. In 1980, this was achieved by taking advantage of a previously created protocol to study Pavlovian conditioning, and applying it in instrumental conditioning (Adams, 1980). Since the mechanistic theory (Descartes et al., 1985; Hull, 1943; Thorndike, 1898) implies an independence of the action from the expected value of the future outcome, if one changes this value, the response should remain the same. However, if the relationship between the action and the outcome (and information about the outcome itself) is encoded during learning, then the animals should be sensitive to changes to the value of the outcome. Thus, animals were trained to press a lever on a variable time (VI) schedule to receive sucrose pellets, which was followed by taste-aversion conditioning of the sucrose pellets. By pairing the reinforcer with a nauseating agent, the value of the outcome was changed. Although it was shown that devaluation of the outcome was effective (and animals would refrain from consuming it), no differences in levels of pressing were observed in a brief test after the outcome devaluation (performed in extinction, where the animal was allowed to press the lever, but it led to no reinforcer delivery). The test was

performed after the outcome devaluation and under extinction to prevent the influence of exposure to the aversive outcome, which would then update the association between action and outcome. This result was, therefore, consistent with the hypothesis that stimulus-response associations, independently of reinforcement encoding, mediate instrumental learning.

Luckily, the behaviour probing did not stop there. Due to how it was designed, the previously described experiments could still have had a context dependent mediated effect. Therefore, a task was developed to directly test if the knowledge of a particular reinforcer was encoded in the learning association (Adams and Dickinson, 1981). Rats were trained to press a lever in a variable ratio (VR) schedule to receive a food pellet, alternated with free delivery of another type of food pellet. To test the sensitivity of the animals to changes in the value of the outcome, either the reinforcing pellets (the paired group) or the free-delivery pellet type (unpaired group) were devalued. It was observed that this protocol led to a change in response of the paired group (compared to the unpaired one) during an extinction test. There was finally evidence that knowledge about the reinforcer could be incorporated in action training in an instrumental task.

It also seemed clear, from these early experiments, that both inflexible stimulus-response behaviour and a more flexible, goal-directed action behaviour can be responsible for instrumental learning. However, the conditions in which each would control the behaviour were not clear, nor how the two action strategies relate to each other. Conceivably, a variety of factors would come into play to determine the extent of goal encoding during learning and performance of an action. One obvious instance of a possible important factor for this balance is repetition of an action. To test

the potential role of repetition in instrumental learning, rats were trained to press a lever to receive a food reinforcer, for either 100 or 500 responses (Adams, 1982). It revealed that not only were moderately trained animals sensitive to outcome devaluation (as seen before), but that continuation of training led to a behaviour autonomous from the current value of the outcome, hence with the properties of stimulus-response learning. Although the aversion protocol used was not reversible, and therefore the observation came from a between subject comparison, the likely conclusion is that for this particular type of instrumental learning, the animals begin the instrumental performance encoding the knowledge about the consequence of that action, and through repetition that action becomes less dependent from the expected value of the outcome, rendering behaviour inflexible. Exactly like what happen to me in my elevator adventures.

One crucial difference in these initial results, which came to be of great value to further probing these questions in a laboratory setting, was the type of schedule utilized. Studies using different reinforcing schedules led to behavioural differences in the sensitivity to changes in the value of the outcome. To make sense of these results, animals were trained under two different schedules of reinforcer: random ratio (RR) - where there is a probability of each response yielding a reinforcement delivery - and random interval (RI) – where the probability of availability of reinforcement to be delivered upon lever press was set in a time dependent manner (Dickinson et al., 1983). It was observed that while animals trained under a RR schedule reduced their responses upon devaluation of the outcome, RI trained animals were insensitive to such changes.

This difference between RR and RI training is now the foundation to several studies which wish to probe and better understand the dichotomy

between goal-directed action and inflexible, stimulus-response, habitual behaviour. Further advances in the way we study these were made by introducing a reversible, within subject test to evaluate devaluation effect during training (Colwill and Rescorla, 1985). After training, the outcome was devalued by satiety to the outcome, allowing the animals "free" consumption of one particular reinforcer, until they refrain from eating it. The subjects are then tested in extinction and their changes in behaviour can be evaluated.

It has become generally accepted that goal-directed actions are characterized as having two main features: there is a representation of the contingency between the action and the outcome; and there is knowledge about the outcome itself (Colwill and Rescorla, 1986; Dickinson and Balleine, 1994). Although the described experiments have focused more on the presence of an outcome representation, there was also evidence for the establishment of an action outcome association that is dependent on their contingency (Colwill and Rescorla, 1986; Dickinson and Charnock, 1985; Hammond, 1980). Although these studies showed differences in their protocol, they all relied on the concept of using free delivery of the reinforcer the animals have been trained to earn. This manipulation led to a degradation of the contingency between the action and the reinforcing, maintaining its previous contiguity. They all observed a decrease in performance of the action in which the contingency was degraded, which also became a characteristic of a goal-directed action.

Since these first observation were made, a substantial amount of information has been gathered about these two behaviour modes, and it is now widely accepted they both coexist in instrumental learning. In the next chapter of this introduction, I will go into more detail on what is known about the circuits mediating both types of action learning. Several

of those were unveiled with the use of lesions, inactivation and electrophysiological recordings in the brain. A further step in the way goal-directed actions and habits are studied was the adaptation of the previously mentioned behavioural paradigms to mice, where scientists can take advantage of the power of genetic manipulation. It was observed, as expected from previous experiments in rats, that mice trained under the RR schedule were sensitive to devaluation, while animals trained under the RI schedule became insensitive to it, pressing equally in both the valued and the devalued sessions (Hilario et al., 2007). Another consequence from the general concept of a habit was also unveiled. If a stimulus-response association depends on previous stimuli eliciting a particular action, rather than the knowledge of its consequence, presentation of similar but novel stimulus could lead to differences in responses for goal-directed and habitual animals. Indeed, when presented with a choice between the training and a novel lever, animals trained under a RR schedule exploited the training lever – a sign of their representation of the contingency learned – while animals trained with a RI schedule generalized to the novel lever, pressing both at a similar rate (Hilario et al., 2007, 2012). A possible (and likely) explanation for this generalization is that the action is under the control of stimuli, which are of a similar character for both levers. As you may remember, generalization of the learned action also happened to me in the elevator.

A major drawback from previous studies was that the comparison of habitual and goal-directed actions was done between animals. This limits the way manipulations and observations can be conducted. It is not possible, for instance, to record simultaneously from the same neuron while animals are performing both types of actions. This has been recently overcome by the development of a task where animals perform the same

action (pressing a lever) to receive the same outcome, but in different contexts (Gremel and Costa, 2013). In one context they were trained under RI, while in the other the training was performed under RR. It was observed that animals readily switch strategies, depending on the context. This training allows for a within animal comparison when probing the system. One can, for instance, lesion an area of the brain and see in the same animal, how that influences both behaviours. Or one can, as mentioned above, record from the same neuron while the same animal is performing a goal-directed versus a habitual action on a lever.

While a lot of speculation as been made on why the different schedules yield different behaviour strategy outputs, this was mostly attributed to the different contingencies the animal was being exposed to (Dickinson, 1985). However, schedules that are controlled for the same level of contingency, making use of the same interval schedules, but with different probabilities of release, raises the intriguing possibility that the observed differences in sensitivity to devaluation and degradation might be due to different experienced contiguity, caused by different uncertainties of the training paradigms (Derusso et al., 2010).

Goal-directed action, habits and the brain

The observation that both action-outcome and stimulus-response learning can occur led to the question of how these are controlled (and represented) in the brain. A growing amount of research has provided evidence that these two types of behaviour depend on dissociable circuits that include different regions of several brain areas, such as prefrontal cortex, amygdala and striatum. This chapter focuses on some of the main advances in

research regarding the areas, as well as the mechanisms, in the brain that control and regulate both behavioural modes.

Of key importance to both goal-directed and habitual behaviours is the basal ganglia, a structure composed of a group of nuclei situated in the cerebrum and composed of: dorsal striatum, nucleus accumbens (also known as ventral striatum), globus pallidus (GP), subthalamic nucleus (STN), substantia nigra (SN) and ventral pallidum (Nelson and Kreitzer, 2014). The globus pallidus is composed of the internal GP (GPi, or entopeduncular nucleus in rodents) and the external GP (GPe). The substantia nigra can be further subdivided in pars compacta (SNc) and pars reticulata (SNr). The striatum is the major input station of the basal ganglia, receiving projects from cortical, thalamic and limbic areas. The PGi and the SNr project to the thalamic nucleus, and constitute the major basal ganglia output pathway. The thalamus, in turn, projects to cortical areas, closing what is considered the striatal-corticothalamic loop (**Figure 1.1**).

The striatum and its subregions on parallel action strategies

In rodents the dorsal striatum is composed of the dorsomedial (DMS, or associative) striatum and the dorsolateral (DLS, or sensorimotor) striatum (Voorn et al., 2004). In primates the equivalent of these two areas are roughly the caudate nucleus and the putamen (Haber, 2003), respectively. While in primates the two structures present a clear anatomic distinction, in the rodent brain there is no demarked separation between the 2 areas. Despite that, it is possible to distinguish them based on protein expression profiles, and functional and anatomical connectivity.

Early studies with lesions have pinpointed the importance of the different areas of the striatum in instrumental learning: Lesioning DLS rendered animals sensitive to outcome devaluation, even under RI training (Hilario et al., 2012; Yin et al., 2004); while lesions in DMS (in particular the posterior DMS) led to animals behaving in an habitual manner (Hilario et al., 2012; Yin et al., 2005a), even in a RR training. It was also observed that DLS is necessary for action generalization, which could be a further hallmark of habitual actions (Hilario et al., 2012).

One amazing observation from these early studies came from post training lesions and inactivations: those experiments could mimic the pre-training lesion results. When inactivating DLS, animals behave in a goal-directed manner (Yin et al., 2006). On the other hand, post-training lesion or inactivation of DMS rendered the animals' behaviour habitual (Yin et al., 2005a). Two main conclusions can be drawn from these results. One is that the areas mentioned are necessary not only for acquisition, but also for expression of each type of behaviour strategy. The second is that both strategies must work in parallel loops, possibly interacting and/or competing with each other, since abolition of one of the control modes does not lead to an inability to perform the action, but just a shift in "why" the animal is performing it.

All this begs the question, what is happening in DMS and DLS during the learning of a new action? What type of plasticity and changes are happening to the neuronal circuits in those areas alongside learning? One first hint came from a study where blockage of NMDA receptor in DMS during the pairing of a known action with a novel reinforcer rendered animals insensitive to changes in the value of that novel reinforcer (Yin et al., 2005b). Since NMDA receptor is associated with long-term plasticity, one can speculate that plasticity in DMS is necessary for goal-directed,

action-outcome associations. Some conclusions about striatal contribution can also be taken from other forms of skill learning. Learning a new skill is seen as a trial and error phase where there is a marked improvement in performance. However, with repetition, the action becomes more like a stereotyped and automatized skill. During learning of a new skilled action, it was observed that although neurons in DMS increased their rate of modulation during the early stages of skill training, their modulation returned to naïve levels later on (Yin et al., 2009). Contrarily, DLS neurons increased their modulation rate in the later stages of skill training. It was also shown that lesions in these areas during training mimicked the recording results. While lesioning DMS impaired performance early in training, but had no effect in late performance, lesions in DLS impaired performance both early and late in training. These were substantiated by ex-vivo recordings of animals in early or later stages of skill learning, where they observed potentiation of synaptic strength in DMS, but only for early stages of training, in contrast to long-term potentiation of DLS neurons with extensive training.

Although with this study we can begin to understand how striatal circuits behave during action learning, it still leaves the question of how they work during habitual learning. And, in particular, we don't know what happens to the same neurons when the animal is performing a goal-directed action versus a habitual one. However, that could be done with the novel 2 context – 2 schedule task (Gremel and Costa, 2013). Using this novel task, where animals can switch between goal-directed or habitual action using contextual cues, it is possible to investigate the changes in neuronal activity between several brain areas. In particular, it was observed that several neurons in both areas were involved in both types of actions, but were modulated differently, depending on devaluation state and brain

area. Analysing the modulation of these neurons suggests that in RR there is a stronger modulation of DMS neurons and a weaker modulation of DLS neurons, compared to RI.

Striatal main subpopulations, action and instrumental learning

The majority of striatal cell (~95%) are GABAergic medium spiny neurons (MSNs) (Kemp and Powell, 1971). The rest of the cells in the striatum are interneurons of several types, including parvalbumin and acetylcholine interneurons. Striatum neurons are usually driven by glutamatergic projection from several cortical areas. As MSNs are usually kept under a very negative membrane potential, and hence distant from their firing threshold, a substantial convergence of cortical input is necessary to drive their activity (Cowan and Wilson, 1994; Wilson, 1993). Although MSNs are similar in their morphology and eletrophisiologycal properties, they are not a homogenous population of neurons. Two major types of MSNs can be found in striatum, which can be divided (and named) based on their projection targets (Kawaguchi et al., 1990): The direct pathway MSNs, also known as striatonigral, directly reach the basal ganglia output nuclei, the GPi/SNr. The second major neuronal population, constituted by the indirect pathway MSNs – or strialpallidal – projects to the GPe. In turn, the GPe projects to the glutamatergic neurons of the STN, which project to the GPi/SNr. Thus, the striatonigral pathway projects directly to the output nucleus of the basal ganglia, while the striatopallidal neurons form an indirect, multisynaptic pathway to the output structures of the basal ganglia (**Figure 1.1**).

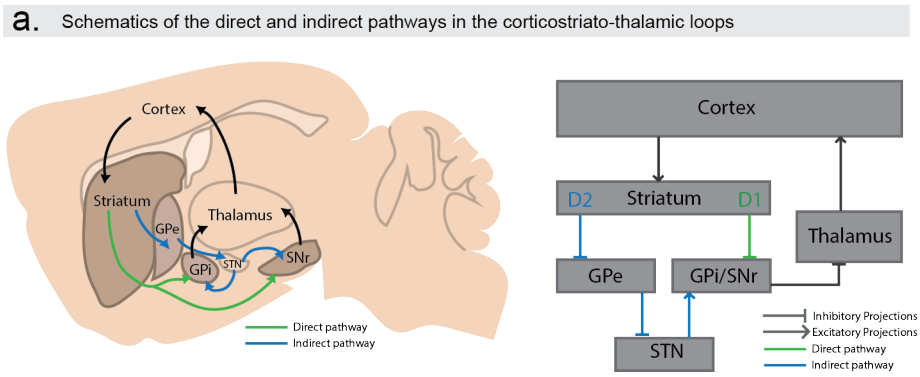

Figure 1.1 | Schematics of both direct and the indirect pathways in the corticostriato-thalamic loops. Projections from neurons of the direct pathway reach the GPi/SNr nuclei directly, while neurons of the indirect pathway communicate to the basal ganglia output nuclei through a multisynaptic connection.

Traditionally the two primary projection pathways of the dorsal striatum are seen as having opposing/antagonistic effects on movement (Albin et al., 1989; Frank et al., 2004). Because of their pattern of projections, a model of functioning of the striatum arose, which stated that the direct pathway was responsible for movement (MSNs inhibit Gpi/SNr, which leads to disinhibition of the thalamus, **Figure 1.2b**) while the indirect pathway leads to cessation of movement (MSNs inhibit GPe, which in turns leads to disinhibition of the STN, which excites GPi/SNr, leading to overall thalamus inhibition, **Figure 1.2b**).

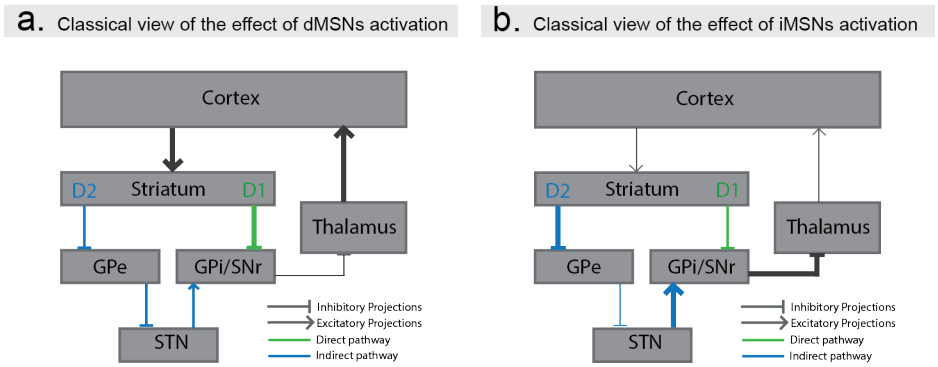

Figure 1.2 | Opposing effects of neurons from the direct and the indirect pathways on thalamic and cortical activity. (a) dMSNs activation leads to inhibition of the basal ganglia output nucleus, which decreases its inhibitory effect in thalamus. **(b)** Conversely, iMSNs activity inhibits GPe, which results in increased activation of GPi/SNr, and thus thalamic inhibition.

This model stipulates that the thalamic-cortical projections are under the control of tonic inhibition from the basal ganglia output. When the directed pathway is activated, this tonic inhibition is lifted, allowing movement. The general opposing role of the indirect pathway is to suppress movement. This dichotomy between pathways has been used, for example, to explain some movement characteristics of Parkinson's disease (PD), where there is degeneration of DA fibres projecting to striatum. It was hypothesized (Albin et al., 1989) that the degeneration creates an imbalance between the two pathways, by changing their excitability in opposing ways.

Importantly, these two MSN populations can also be distinguished based on their expression of dopamine (DA) receptors. Striatonigral neurons selectively express the D1 DA receptor, while striatopallidal neurons express the D2 DA receptor (Gerfen et al., 1990). DA modulates the processing of input from cortical and thalamic areas onto MSNs, and a general model has emerged regarding the different way this DA

modulation happens in striatonigral and striatopallidal neurons. Previous studies have shown that DA interacts differently with both types of receptor. D1 receptor has low DA affinity, while D2 receptor presents high affinity for DA (Richfield et al., 1989). Consequently, low DA states (like DA tonic release) should lead to activation of D2-receptors, while high concentrations of DA (such as in dopamine burst) should activate D1-receptors. In the classical model of DA dependent modulation of activity, activation of D1 receptors leads to an excitation state of the striatonigral neurons, while activation of D2 receptors reduces the excitability of striatopallidal neurons (Surmeier et al., 2007). Interestingly, the effect on increased excitability of striatonigral neurons upon D1 activity seems to necessitate strong and coordinated glutamatergic input, rather than just leading to a general increase in the excitability of the cell (reviewed in Surmeier et al., 2007). The effects on both cell types depend on activations of cascades in the cells, which ultimately lead to changes in ion channels that regulate how excitable the cell is.

It has been suggested that striatonigral and striatopallidal neurons have opposing roles in not only action performance, but also action reinforcement (Albin et al., 1989; DeLong, 1990; Frank et al., 2004). As mentioned before, the activity of the striatonigral neurons is traditionally associated with the generation of movements, due to a disinhibition of the thalamus, while the striatopallidal pathway is seen as inhibiting movement. Concomitant with their role action activation/inhibition, the dichotomy in these two pathways seems to extend to reinforcement. One can easily use the classical model idea to think of a simple mechanism by which this is at least possible. It is known that a DA release is associated with the delivery of unexpected reward. Since transient DA increases excitability of striatonigral neurons but decreases excitability of striatopallidal neurons,

one could conjure that positive reinforcement will lead to activation of the direct pathway, which will in turn lead to action performance, while decreasing the activity of the indirect pathway (thus decreasing action suppression). Accordingly, studies have shown striatonigral neurons as being important to learn positive reinforcement and indirect pathway neurons as being important to learn to avoid undesired actions (Go/No-Go) (Frank et al., 2004). It has also been shown that optogenetic self-stimulation of striatonigral neurons in DMS leads to reinforcement of actions that lead to stimulation, while self-stimulation of striatonigral neurons lead to avoidance of actions that lead to stimulation (Kravitz et al., 2012).

An alternative explanative functional model of the basal ganglia was proposed by Mink (Mink, 2003). In this model both striatal projection pathways are involved in action selection, with striatonigral neurons supporting the execution of the desired action pattern, and striatopallidal neurons avoiding the execution of competing actions. Consistent with this theory, studies have observed a transient increase in neuronal activity in both direct and indirect pathways during action initiation (Cui et al., 2013). Furthermore, in a sequence lever pressing task, both striatopallidal and striatonigral neurons present activity associated with the start and the end of a lever pressing sequence (Jin et al., 2014). Finally, it has been shown that in non-instrumental movement, both D1 and D2 neurons show an increase in activity during spontaneous contralateral movements (Tecuapetla et al., 2014). Inhibition of either or both pathways was also shown to impair contralateral movements. Taken together, these studies provided considerable evidence to the hypothesis that co-activation of both striatopallidal and striatonigral is necessary for movement and action selection.

Bringing all these pieces of evidence together, and to further explore these distinct subcircuits, one can now ask if within the subregions of dorsal striatum - DMS and DLS - both pathways have the same behaviour during instrumental learning. One of the first factors to pay attention to would be the comparative distribution of D1 and D2 receptors in DMS versus DLS. In both rats and mice, D2 receptor expression seems to be more abundant in DLS, compared to DMS (Joyce et al., 1985; Yin et al., 2009). While the distribution of D1 might be more dependent on the species (even within rodents), in mice its expression appears evenly distributed between both striatal subregions (Savasta et al., 1986; Yin et al., 2009).

In skill learning, both inactivation of D1 or D2 receptors affected early skill learning. With extensive training, however, performance becomes less dependent on D1 receptor activation, but still dependent on D2 receptor activation (Yin et al., 2009). Looking at the potentiation of the different subpopulation of neurons with training, specifically in DLS (which is necessary for late skill learning), it was observed that both pathways showed potentiation of synaptic strength, but that this potentiation was more prominent in iMSNs (Yin et al., 2009). Taken together, these observations point to a possible role in late learning for striatopallidal neurons in DLS. Corroborating this premise, and bringing the topic back to the balance between goal-directed actions and habits, it has been shown that striatal-specific deletion of A_{2A}-receptor, which is specifically expressed in striatopallidal neurons and abolishes long-term potentiation onto those indirect MSNs, impairs habit formation (Yu et al., 2009). These data suggest that, in DLS, striatopallidal MSNs are involved in action learning and positive reinforcement, but could possibly support a different action strategy from striatonigral neurons. In chapter 4 we investigate the role of dMSNs and iMSNs in DLS in action reinforcement.

Glutamatergic inputs to dorsal striatum, goal-directed actions and habits

As mentioned before, the striatum is considered the major input station of the basal ganglia. As such, it receives extensive projection from various brain regions, including cortical, thalamic and amygdala areas. Interestingly, most of those projections present a functional and topographic organization that closely relates to the already mentioned functional segregated areas of striatum.

Thalamic projections to the striatum (and to the dorsal part of the striatum in particular) follow a gradient in the medial-lateral axis. While the posterior and lateral intralaminar thalamic nuclei project to the DLS areas, more medial intralaminar thalamic nuclei project to more medial dorsal striatum (Haber, 2003; Voorn et al., 2004). In accordance with the projections, animals with lesions in the mediodorsal thalamus are insensitive to changes in the value of the outcome (Corbit et al., 2003).

Cortical afferents are one of the main glutamatergic drivers of activity in the striatum. The cortical projections to striatum, as in the case of thalamus, reflect the medial-lateral gradient of dorsal striatum. In particular, DMS receives cortical inputs from associative areas of the cortex, while DLS receives inputs from more sensorimotor involved areas of the cortex (Berendse et al., 1992; Haber, 2003; McGeorge and Faull, 1989; Voorn et al., 2004). Different cortical areas have also been implicated in learning and/or expressing of goal-directed and habitual action. Pre-training lesions of prelimbic cortex, which projects to DMS (Vertes, 2004), render animals insensitive to devaluation of the outcome (Killcross and Coutureau, 2003). Post-training lesions do not, however, yield the same result, rendering animals with intact sensitivity to outcome devaluation (Ostlund, 2005). On the other hand, both pre- and post-

training inactivation of infralimbic cortex led to animals that do not form habits, i.e. are sensitive to changes in the value of the outcome even after extensive training (Coutureau and Killcross, 2003; Killcross and Coutureau, 2003). OFC (which projects to dorsal striatum (Schilman et al., 2008)) lesioned animals are also insensitive to devaluation of the outcome (Gremel and Costa, 2013). Recordings from OFC in the same animals performing a task under RR or RI schedules show that modulation of neurons in this area during pressing RR versus RI in de devalued state have a positive correlation with degree of goal-directedness. Further, while activation of OFC during devaluation testing after RI training had no effect on pressing behaviour, activation particularly during the devalued state after RR training increased the number of lever presses. Taken together, these results point to the important role of OFC in valuing action in striatum. However, besides striatum, OFC projects to other areas associated with goal-directed expression of behaviour, such as basolateral amygdala (BLA) and ventral tegmental area (VTA) (Gremel and Costa, 2013), which leads to the question of whether the value association is made directly through the projection of OFC to striatum, or through its projection with other areas. Of particular interest to this question is the amygdala, whose role has been less explored in instrumental learning, and which will be mentioned below.

The amygdala is a deep midbrain structure composed of several anatomical and physiological distinct nuclei that partake in the limbic system of the brain. The exact number of nuclei or subareas that constitute the amygdala is constantly under debate. Among these nuclei are the basolateral amygdala (BLA, which can be further subdivided in lateral, basal and basomedial amygdala), and the central nucleus (CeA, which can be further divided in lateral and medial central amygdala) (Krettek and

Price, 1978). The BLA is a cortical-like structure, formed by a large majority (around 80%) of glutamatergic pyramidal neurons and a net of heterogeneous inhibitory aspiny GABAergic interneuron population (Millhouse and DeOlmos, 1983). CeA is a structure that resembles the striatum, constituted by a majority of GABAergic projection neurons (McDonald, 1982). Amygdala is in general silent and kept under tonic inhibition. Measured pyramidal cells in rats have a mean resting potential of -69.5 and do not spontaneously firing at membrane resting potential (Washburn and Moises, 1992).

Amygdala receives input from all sensory input modalities, and sends projections to several brain areas (Sah et al., 2003). BLA – and its lateral subdivision in particular – is traditionally considered the sensory intput of the amygdala. Amygdala has several intra- and inter- connections between its nuclei. Some amygdala nuclei also have projections to the contralateral amygdala. BLA projects broadly to striatum. Although these projections are broad, very interestingly they mostly avoid DLS, in particular the most antero-DLS (Kelley et al., 1982). With the help of a 2 virus system that maps monosynaptic connections, Wall et al have also observed that amygdala relayed input mostly to D1 neurons in the dorsal striatum (Wall et al., 2013).

Amygdala has been described as important in a variety of behaviours related with emotional behaviour, such as fear, anxiety and stress, as well as in appetitive and aversive conditioning (Balleine and Killcross, 2006; Grundemann and Luthi, 2015). The first ideas of amygdala functions came from lesion studies in primates (Brown and Schäfer, 1888). Investigators observed that lesions around this area (but also encompassing other brain structures) led to animals with intact sensory perception, but impaired responses to said stimuli and reduced fear displays. Furthermore, in 1956,

Weiskrantz observed that lesions in amygdala tended to impair reinforcing stimuli identification (Weiskrantz, 1956). Since then, several studies on rodents (mainly using sound and shock associations – US and CS associations) have identified the amygdala as a key structure in fear expression and fear conditioning (for a review on this literature, please refer to Duvarci and Pare, 2014). Lesions and inactivations of both BLA and CeA impair acquisition and/or expression of conditioned fear responses.

Simultaneously, a growing amount of evidence has been collected regarding the function of amygdala, BLA in particular, in associations between neutral stimulus and natural rewards (such as food or sexual behaviour). It was also observed that BLA is necessary for conditioned place preference (CPP) normal expression, and that functional disconnections between BLA and ventral striatum disrupted CPP (Everitt et al., 1991). Hatfield and colleagues (1996) have also shown that BLA is necessary for acquisition of a second order Pavlovian conditioning (Hatfield et al., 1996). It has also shown that although lesions in BLA did not impair US-CS conditioning, post-train devaluation of the US with LiCl did not diminished conditional responses to the CS in lesioned animals (Hatfield et al., 1996). Therefore, BLA appeared to be involved in adjustment of responses to changes in the value of the US. The importance of BLA during devaluation testing was also asserted by Malkova et al. (1997), where amygdala-lesioned monkeys that were trained to discriminate and choose between visual stimuli could performed the task, but were insensitive to devaluation by specific satiety (Málková et al., 1997). However, the question of how the neurons in amygdala, and in particular BLA, respond to conditioning and changes in value was still largely unanswered. In rats trained in an odour discrimination task, where

odour predicted the delivery of a positive or negative reinforcer (Schoenbaum et al., 1998), it was observed that cells in BLA, during early training, began responding to the delay period before reinforcer delivery. These signals were consistent with BLA's involved in the association of a response and the expected value of its outcome. Neurons in BLA also developed early selectivity to a stimulus associated with a specific reinforcer valence; that selectivity was reversed if the valence of the reinforcer associated with that specific odour changed. In monkeys it was also observed that amygdala neurons encoded the value of images associated with positive or negative USs, in separate neurons, and in a flexible manner, that accompanied and correlated with changes in learning the value of such visual stimuli (Paton et al., 2006).

In cats, physiological recordings suggest that there is an increase in coupling of gamma waves between BLA and striatum during learning (Popescu et al., 2009). They also observed that disruption in BLA activity reduced gamma activity in striatum, and that states of high striatal gamma correlated with increased coupling of BLA and striatal neuronal activity.

One BLA characteristic that is easily perceived from early (and recent studies) is its bivalence regarding valence encoding. To further analyse distinct positive and negative responding populations, Gore *et al* (2015) use a virus expressing both mCherry and ChR under the control of c-fos (Gore et al., 2015). With this technique, they could confirm that different intermingled neuronal populations in BLA were active after either positive or negative US. Additionally, pairing stimulation of the negative BLA cells (i.e. marked with the negative reinforcer) and a tone led to increase freezing towards that tone. Reversely, pairing stimulation of positive BLA cells (i.e. marked with the positive reinforcer) with an odour led to an increase in approaching behaviour towards the odour. Finally, they

observed that after learning there was a co-localization between neurons that respond to US and to CS, which is an indication that learning takes place through the association of US responding cells with the CS, and that US responding neurons are necessary for the expression of the learned behaviour towards the CS.

Since BLA has intermingled populations of neurons that code positive and negative values, one could speculate on how they give rise to very different behaviour outputs (e.g., stimuli approaching or freezing). In a set of cleverly designed experiments, Namburi *et al* (2015) looked at different BLA neuronal populations regarding their projection targets (Namburi et al., 2015). In particular, they observed that in BLA to NAc projecting neurons, AMPAR/NMDAR ratio increased after pairing of sucrose with a CS, while decreasing during pairing of a CS with a foot shock. On the contrary, BLA to medial CeA projecting neurons had the opposite modulation of AMPAR/NMDAR ratio. Furthermore, pairing BLA-NAc neuronal stimulation with ChR upon nose-poking led to an increase in that behaviour, while pairing BLA-CeA stimulation led to place avoidance.

Several studies have, therefore, pinpointed the importance of amygdala in tracking the value of stimuli, and in pavlovian conditioning of previously neutral stimuli with specific outcomes. But the question about the role of BLA (and amygdala in general) in instrumental learning remained. Can BLA be the source of value information necessary to learn a new action? How does activity in BLA cells change during action learning? It was clear from earlier described experiments – such as satiety specific devaluation – that when a new action-outcome is formed, the learned association is dependent on the value of the reinforcer. In a two actions – two outcomes task, it was observed that animals with lesions in BLA were able to acquire an instrumental task with no visible impairment, but were insensitive to

devaluation in a choice test (Balleine et al., 2003). These animals were also insensitive to degradation of the action-outcome contingency. An important observation to take out of this study is that operant learning per se was not affected – or at least its behaviour output was intact. Therefore, although BLA appears to be necessary for value assignment to a specific action-outcome, it is not necessary to engage general valuation (or reinforcing) mechanisms. In this work authors suggest that the capacity to acquire the behaviour, in the absence of BLA, might be driven by brain circuits responsible for habitual action learning. Furthermore, and an important detail in both studies, although animals with BLA lesions were insensitive to devaluation in extinction, when they were tested with reinforcer delivery, they partially recovered sensitivity along the testing session. It is also relevant that, in a simple feeding test after devaluation, animals accordingly adjusted consumption of the devalued and valued reinforcer, which indicates that the capacity to discriminate different outcomes was intact. Post-training BLA lesions also disrupt sensitivity to outcome devaluation (Ostlund and Balleine, 2008).

Lesions in the anterior CeA, on the other hand, rendered animals more sensitive to outcome devaluation, even in a training schedule that rendered control sham animals habitual (Corbit and Balleine, 2005)(Lingawi and Balleine, 2012). Functional disconnections between CeA and DLS, necessary for habitual action performance, also led to animals with increased sensitivity to outcome devaluation (Lingawi and Balleine, 2012). Disconnections between BLA and DMS, as expected from previous results, resulted in animals that were insensitive to devaluation (Corbit et al., 2013).

Taken together, this set of experiments exposes a possible dissociable and parallel network in amygdala, similar to that of dorsal striatum and cortex,

responsible for different behaviour strategies of instrumental learning. It also appears evident that amygdala is a key player in updating the value of an US, and associating changes of value of the US with behaviour. It is not clear, however, how these areas behave and interact during instrumental learning; or if the interacting between BLA and CeA, and DMS and DLS, is controlled by direct connections between those areas or indirect connections (passing, for example, through orbitofrontal cortex or substantia nigra). Interestingly, it was also recently observed that inactivation of BLA projecting to CeA neurons not only led to a decrease in freezing, but also in an increase in conditioned reward seeking (Namburi et al., 2015). Therefore, it is also possible that this system is not only regulating/modulating behaviour through their down-stream targets, but also that direct intra-amygdalar projection play a role in this balance. In chapter 3 we investigate the role of BLA to DMS projections in learning of a novel instrumental action.

Dopamine modulation of dorsostriatal circuits involved in habitual and goal-directed actions

Dopamine is a monoamine neuromodulator present in many eukaryotes, with an important role in movement, aggression, sexual behavior, reward and learning. This neurotransmitter was first identified in 1957 (Carlsson et al., 1957, 1958). The role of dopamine in different stages of learning in instrumental conditioning is complex, since DA signalling is regulated by several factors, including different mechanisms of release, re-uptake and molecular expression of receptors. The two main sources of DA in the brain are the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc) (Fallon and Moore, 1978). DA neurons projecting from

these areas form dense arbours into several brain areas. Many of such DA neurons collateralize to multiple regions. Dopamine neurons have two modes of activity: they can fire in a tonic pacemaker-firing pattern (2-8 spikes/sec in rat and mice) or in transient synchronous bursts (sharp increases/decreases of firing rate for about 100-500 ms, with an iter-spike interval of about 70ms). DA autoreceptores present in DA neurons provide feedback, regulating DA release by decreasing DA synthesis and increasing its reuptake (Schmitz, 2013). DA VTA neurons have also been shown to co-release glutamate (which is not observed in SNc DA neurons) (Koos et al., 2011).

A major advance in DA and reinforcement learning was the discovery of DA phasic burst activity upon unexpected rewards, which progresses to responses to unexpected reward-predicting stimuli (Fiorillo et al., 2003; Schultz, 1998). This was interpreted as a reward prediction error, which could be considered a teaching signal that allows learning. It is worth noting however, that DA neurons can also respond with bursting signals to salient, non-rewarded sensory stimuli (Bromberg-Martin et al., 2011; Horvitz, 2000).

One of the areas strongly innervated by DA neurons is the striatum. In the striatum, axons from one single DA cell branch enormously and form dense arborizations. Dopamine is involved in bidirectional, time-dependent corticostriatal LTP (long-term potentiation) and LTD (long-term depression) plasticity in both D1 and D2 MSNs (Shen et al., 2008). In D1 MSNs, bidirectionality of spike-timing-dependent plasticity is dependent upon D1 receptor activity, such that positive timing that normally results in LTP will lead to LTD under D1 antagonist exposure. In D2 MSNs, positive timing will lead to LTP and negative timing to LTD (as usual), but LTD is abolished with D2 antagonist exposure and further reverted to LTP with

A2a-receptor agonists. Therefore, dopamine is essential for the bi-directionality of the plasticity in the different MSN populations. DA transmission has also been linked with dynamically regulating the coordination of activity between cortex and striatum – while low DA levels are associated with states of high correlation, during hyperkinesia and high DA transmission there are lower levels of cross-correlation between neurons in cortex and striatum (Costa et al., 2006). Linking all together, one current theory of DA functioning asserts its necessity in reinforcement learning by its involvement in adjusting the strength of synaptic connections between neurons (Bromberg-Martin et al., 2011).

Although dopamine neurons project broadly to striatum, they present a topographic distribution to the different areas of dorsal striatum; while DMS receives projections from VTA and the ventral medial SNc, DLS is innervated from the dorsolateral SNc (Fallon and Moore, 1978). DA depletion in different striatal subregions is also distinctively associated with different behavioural modes. DA depletion specifically in pDMS led to animals insensitive to contingency degradation, while maintaining intact their sensitivity to outcome devaluation. Faure and collegues (Faure et al., 2005) have also observed that animals with DA lesions in DLS were sensitive to contingency devaluation, even with overtraining. Endocannabinoid signaling, which is critical for habit formation (Hilario et al., 2007), is dependent upon dopamine signaling in dorsal striatum and shows a gradient expression in dorsal striatum, with the maximum expression in DLS.

Another striking difference regarding DA signalling in different areas of dorsal striatum is the mechanism of DA clearance from the synapse. In DMS there is a prevalence of catechol-O-methyltransferase (COMT) (Matsumoto et al., 2003), which degrades dopamine in the synapse.

However, the dopamine transporter (DAT), which re-uptakes DA back to the cell, has an increasing gradient of expression from DMS to DLS. The main source of this difference is the differential expression of DAT in the midbrain DA neurons, since DAT is expressed more in SNc and less in VTA (Blanchard et al., 1994). Within DA neurons' axons, DAT is mainly located outside synaptic active zones (Hersch et al., 1997), and there are studies implying a role of DAT in direct control of DA signalling (Gowrishankar et al., 2014).

DAT is a transmembrane transporter with 12 domains, with a carboxyl and an amino cytoplasmic terminal (German et al., 2015). DAT is the major target for some psychostimulants, such as cocaine and amphetamine (AMPH). AMPH, for example, has been shown to interfere with DAT functioning in more than one way; this psycostimulant induces an efflux of DA via DAT (inverting its functioning), competes with DA binding to the transporter and induces DAT internalization (Kahlig et al., 2005). Interestingly, sensitization with amphetamine induces spine changes in MSNs (Jedynak et al., 2007), increasing spines in DLS and decreasing their density in DMS. These appear to mimic the normal changes upon a transition from goal-directed to habitual actions. Behaviourally, and in agreement with the spines physiological changes, sensitization has been shown to accelerate habit formation in rats (Nelson and Killcross, 2006).

Arguably, one of the most studied conditions traditionally associated with the habitual system is obsessive-compulsive disorder (OCD) (Gillan et al., 2015). DAT functioning and its role in this pathology is still under debate, with some studies observing an increase in DAT binding in OCD patients (Kim et al., 2003), while other report lower DAT availability in striatum (Hesse et al., 2005). One study in dogs with obsessive compulsive behaviours further confirmed these results, by observing altered levels of

DAT binding, but in both directions (Vermeire et al., 2012). Furthermore, subjects in early alcohol withdrawal show a hypermethylation (usually associated with a non-transcription state) of the DAT promoter (Hillemacher et al., 2009). Chapter 2 is devoted to further study the role of DAT in the balance between goal-directed actions and habits.

REFERENCES

Adams, C.D. (1980). Post-conditioning devaluation of an instrumental reinforcer has no effect on extinction performance. Q. J. Exp. Psychol. *32*, 447–458.

Adams, C.D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. Q. J. Exp. Psychol. Sect. B *34B*, 77–98.

Adams, C.D., and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. Q. J. Exp. Psychol. Sect. B Comp. Physiol. Psychol. *33*, 109–121.

Albin, R.L., Young, A.B., and Penney, J.B. (1989). The functional anatomy of basal ganglia disorders. Trends Neurosci. *12*, 366–375.

Balleine, B.W., and Killcross, S. (2006). Parallel incentive processing: an integrated view of amygdala function. Trends Neurosci. *29*, 272–279.

Balleine, B.W., Killcross, a S., and Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. J. Neurosci. *23*, 666–675.

Berendse, H.W., Galis-de Graaf, Y., and Groenewegen, H.J. (1992). Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. J. Comp. Neurol. *316*, 314–347.

Blanchard, V., Raisman-Vozari, R., Vyas, S., Michel, P.P., Javoy-Agid, F., Uhl, G., and Agid, Y. (1994). Differential expression of tyrosine hydroxylase and membrane dopamine transporter genes in subpopulations of dopaminergic neurons of the rat mesencephalon. Mol. Brain Res. *22*, 29–38.

Bromberg-Martin, E.S., Matsumoto, M., and Hikosaka, O. (2011). Dopamine in motivational: rewarding, aversive,and alerting. Neuron *68*, 815–834.

Brown, S., and Schäfer, E. (1888). An investigation into the functions of the occipital and temporal lobes of the monkey's brain. Phil. Trans. R. Soc. B *179*, 303–327.

Carlsson, A., Lindqvist, M., and Magnusson, T. (1957). 3,4-Dihydroxyphenylalanine and 5-

hydroxytryptophan as reserpine antagonists. Nature *180*, 1200.

Carlsson, A., Lindqvist, M., Magnusson, T., and Waldeck, B. (1958). On the presence of 3-hydroxytyramine in Brain. Science *127*, 471.

Colwill, R.M., and Rescorla, R.A. (1985). Postconditioning Devaluation of a Reinforcer Affects Instrumental Responding. *I*, 120–132.

Colwill, R.M., and Rescorla, R.A. (1986). Associative Learning Structures in Instrumental Learning. Psychol. Learn. Motiv. *20*, 31–33.

Corbit, L.H., and Balleine, B.W. (2005). Double Dissociation of Basolateral and Central Amygdala Lesions on the General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer. J. Neurosci. *25*, 962–970.

Corbit, L.H., Muir, J.L., and Balleine, B.W. (2003). Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. Eur. J. Neurosci. *18*, 1286–1294.

Corbit, L.H., Leung, B.K., and Balleine, B.W. (2013). The role of the amygdala-striatal pathway in the acquisition and performance of goal-directed instrumental actions. J Neurosci *33*, 17682–17690.

Costa, R.M., Lin, S.-C., Sotnikova, T.D., Cyr, M., Gainetdinov, R.R., Caron, M.G., and Nicolelis, M.A.L. (2006). Rapid Alterations in Corticostriatal Ensemble Coordination during Acute Dopamine-Dependent Motor Dysfunction. Neuron *52*, 359–369.

Coutureau, E., and Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav. Brain Res. *146*, 167–174.

Cowan, R.L., and Wilson, C.J. (1994). Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. J Neurophysiol *71*, 17–31.

Cui, G., Jun, S.B., Jin, X., Pham, M.D., Vogel, S.S., Lovinger, D.M., and Costa, R.M. (2013). Concurrent activation of striatal direct and indirect pathways during action initiation. Nature *494*, 238–242.

DeLong, M.R. (1990). Primate models of movement disorders of basal ganglia origin. Trends Neurosci. *13*, 281–285.

Derusso, A.L., Fan, D., Gupta, J., Shelest, O., Costa, R.M., and Yin, H.H. (2010). Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. Front. Integr. Neurosci. *4*, 1–8.

Descartes, R., and Clark, D.M. (1999). Discourse on Method, and related writings (London: Penguin Classics).

Descartes, R., Cottingham, J., Stoothoff, R., and Murdoch, D. (1985). The philosophical writings of Descartes (Cambridge: Cambridge University press).

Dickinson, A. (1985). Actions and Habits: The Development of Behavioural Autonomy. Philos. Trans. R. Soc. B Biol. Sci. *308*, 67–78.

Dickinson, A., and Balleine, B.W. (1994). Motivational control of goal-directed action. Anim. Learn. Behav. *22*, 1–18.

Dickinson, A., and Charnock, D.J. (1985). Contingency effects with maintained instrumental reinforcement. Q. J. Exp. Psychol. Sect. B *37*, 397–416.

Dickinson, A., Nicholas, D.J., and Adams, C.D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. Q. J. Exp. Psychol. *35B*, 35–51.

Duvarci, S., and Pare, D. (2014). Amygdala Microcircuits Controlling Learned Fear. Neuron *82*, 966–980.

Everitt, B.J., Morris, K.A., O'Brien, A., and Robbins, T.W. (1991). The basolateral amygdala-ventral striatal system and conditioned place preference: Further evidence of limbic-striatal interactions underlying reward-related processes. Neuroscience *42*, 1–18.

Fallon, J.H., and Moore, R.Y. (1978). Catecholamine innervation of the basal forebrain. IV. Topography of the dopamine projection to the basal forebrain and neostriatum. J. Comp. Neurol. *180*, 545–580.

Faure, A., Haberland, U., Conde, F., and El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. J. Neurosci. *25*, 2771–2780.

Fiorillo, C.D., Tobler, P.N., and Schultz, W. (2003). Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. Science (80-.). *299*, 1898–1902.

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

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*, 1429–1432.

German, C.L., Baladi, M.G., Mcfadden, L.M., Hanson, G.R., and Fleckenstein, A.E. (2015). Regulation of the Dopamine and Vesicular Monoamine Transporters : Pharmacological Targets and Implications for Disease. 1005–1024.

Gillan, C.M., Robbins, T.W., Sahakian, B.J., van den Heuvel, O.A., and van Wingen, G. (2015). The role of habit in compulsivity. Eur. Neuropsychopharmacol. 1–13.

Gore, F., Schwartz, E.C., Brangers, B.C., Aladi, S., Stujenske, J.M., Likhtik, E., Russo, M.J., Gordon, J.A., Salzman, C.D., and Axel, R. (2015). Neural Representations of Unconditioned Stimuli in Basolateral Amygdala Mediate Innate and Learned Responses. Cell *162*, 134–145.

Gowrishankar, R., Hahn, M.K., and Blakely, R.D. (2014). Good riddance to dopamine:

Roles for the dopamine transporter in synaptic function and dopamine-associated brain disorders. Neurochem. Int. *73*, 42–48.

Gremel, C.M., and Costa, R.M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nat. Commun. *4*, 2264.

Grundemann, J., and Luthi, A. (2015). Ensemble coding in amygdala circuits for associative learning. Curr. Opin. Neurobiol. *35*, 200–206.

Haber, S.N. (2003). The primate basal ganglia: Parallel and integrative networks. J. Chem. Neuroanat. *26*, 317–330.

Hammond, L.J. (1980). The effect of contingency upon the appetitive conditioning of free-operant behavior. J. Exp. Anal. Behav. *34*, 297–304.

Hatfield, T., Han, J.S., Conley, M., and Gallagher, M. (1996). Neurotoxic Lesions of Basolateral, But Not Central, Amygdala Interfere with Pavlovian Second-Order Conditioning and Reinforcer Devaluation Effects. J. Neurosci. *16*, 5256–5265.

Hersch, S.M., Yi, H., Heilman, C.J., Edwards, R.H., and Levey, A.I. (1997). Subcellular localization and molecular topology of the dopamine transporter in the striatum and substantia nigra. J. Comp. Neurol. *388*, 211–227.

Hesse, S., Müller, U., Lincke, T., Barthel, H., Villmann, T., Angermeyer, M.C., Sabri, O., and Stengler-Wenzke, K. (2005). Serotonin and dopamine transporter imaging in patients with obsessive-compulsive disorder. Psychiatry Res. - Neuroimaging *140*, 63–72.

Hilario, M.R.F., Clouse, E., Yin, H.H., and Costa, R.M. (2007). Endocannabinoid signaling is critical for habit formation. Front. Integr. Neurosci. *1*, 1–12.

Hilario, M.R.F., Holloway, T., Jin, X., and Costa, R.M. (2012). Different dorsal striatum circuits mediate action discrimination and action generalization. Eur. J. Neurosci. *35*, 1105–1114.

Hillemacher, T., Frieling, H., Hartl, T., Wilhelm, J., Kornhuber, J., and Bleich, S. (2009). Promoter specific methylation of the dopamine transporter gene is altered in alcohol dependence and associated with craving. J. Psychiatr. Res. *43*, 388–392.

Horvitz, J.C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non- reward events. Neuroscience *96*, 651–656.

Hull, C.L. (1943). Principles of Behavior: an Introduction to Behavior Theory (New York).

Jedynak, J.P., Uslaner, J.M., Esteban, J. a., and Robinson, T.E. (2007). Methamphetamine-induced structural plasticity in the dorsal striatum. Eur. J. Neurosci. *25*, 847–853.

Jin, X., Tecuapetla, F., and Costa, R.M. (2014). Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. Nat. Neurosci. *17*, 423–430.

Joyce, J.N., Loeschen, S.K., and Marshall, J.F. (1985). Dopamine D-2 receptors in rat

caudate-putamen: The lateral to medial gradient does not correspond to dopaminergic innervation. Brain Res. *338*, 209–218.

Kahlig, K.M., Binda, F., Khoshbouei, H., Blakely, R.D., McMahon, D.G., Javitch, J. a, and Galli, A. (2005). Amphetamine induces dopamine efflux through a dopamine transporter channel. Proc. Natl. Acad. Sci. U. S. A. *102*, 3495–3500.

Kawaguchi, Y., Wilson, C.J., and Emson, P.C. (1990). Projection subtypes of rat neostriatal matrix cells revealed by intracellular injection of biocytin. J. Neurosci. *10*, 3421–3438.

Kelley, A.E., Domesick, V.B., and Nauta, W.J.H. (1982). The amygdalostriatal projection in the rat-an anatomical study by anterograde and retrograde tracing methods. Neuroscience *7*.

Kemp, J.M., and Powell, T.P.S. (1971). The structure of the caudate nucleus of the cat: light and electron microscopy. Philos. Trans. R. Soc. B Biol. Sci. *262*, 383–401.

Killcross, S., and Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb. Cortex *13*, 400–408.

Kim, C.H., Koo, M.S., Cheon, K.A., Ryu, Y.H., Lee, J.D., and Lee, H.S. (2003). Dopamine transporter density of basal ganglia assessed with [123I]IPT SPET in obsessive-compulsive disorder. Eur. J. Nucl. Med. Mol. Imaging *30*, 1637–1643.

Koos, T., Tecuapetla, F., and Tepper, J.M. (2011). Glutamatergic signaling by midbrain dopaminergic neurons: Recent insights from optogenetic, molecular and behavioral studies. Curr. Opin. Neurobiol. *21*, 393–401.

Kravitz, A. V, Tye, L.D., and Kreitzer, A.C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat. Neurosci. *15*, 816–818.

Krettek, J.E., and Price, J.L. (1978). A description of the amygdaloid complex in the rat and cat with observations on intra-amygdaloid axonal connections. J. Comp. Neurol. *178*, 255–280.

Lingawi, N.W., and Balleine, B.W. (2012). Amygdala Central Nucleus Interacts with Dorsolateral Striatum to Regulate the Acquisition of Habits. J. Neurosci. *32*, 1073–1081.

Málková, L., Gaffan, D., and Murray, E.A. (1997). Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. J. Neurosci. *17*, 6011–6020.

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

McDonald, A.J. (1982). Cytoarchitecture of the central amygdaloid nucleus of the rat. J. Comp. Neurol. *208*, 401–418.

McGeorge, A.J., and Faull, R.L.M. (1989). The organization of the projection from the cerebral cortex to the striatum in the rat. Neuroscience *29*, 503–537.

Millhouse, O.E., and DeOlmos, J. (1983). Neuronal configurations in lateral and basolateral amygdala. Neuroscience *10*, 1269–1300.

Mink, J.W. (2003). The Basal Ganglia and Involuntary Movements. Arch Neurol *60*, 1365–1368.

Namburi, P., Beyeler, A., Yorozu, S., Calhoon, G.G., Halbert, S.A., Wichmann, R., Holden, S.S., Mertens, K.L., Anahtar, M., Felix-Ortiz, A.C., et al. (2015). A circuit mechanism for differentiating positive and negative associations. Nature *520*, 675–678.

Nelson, A., and Killcross, S. (2006). Amphetamine exposure enhances habit formation. J. Neurosci. *26*, 3805–3812.

Nelson, A.B., and Kreitzer, A.C. (2014). Reassessing Models of Basal Ganglia Function and Dysfunction. Annu. Rev. Neurosci. *37*, 117–135.

Ostlund, S.B. (2005). Lesions of Medial Prefrontal Cortex Disrupt the Acquisition But Not the Expression of Goal-Directed Learning. J. Neurosci. *25*, 7763–7770.

Ostlund, S.B., and Balleine, B.W. (2008). Differential involvement of the basolateral amygdala and mediodorsal thalamus in instrumental action selection. J. Neurosci. *28*, 4398–4405.

Paton, J.J., Belova, M.A., Morrison, S.E., and Salzman, C.D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature *439*, 865–870.

Popescu, A.T., Popa, D., and Paré, D. (2009). Coherent gamma oscillations couple the amygdala and striatum during learning. Nat. Neurosci. *12*, 801–807.

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*, 767–777.

Sah, P., Faber, E.S.L., Lopez De Armentia, M., and Power, J. (2003). The amygdaloid complex: anatomy and physiology. Physiol. Rev. *83*, 803–834.

Savasta, M., Dubois, A., and Scatton, B. (1986). Autoradiographic localization of D1 dopamine receptors in the rat brain with [3H]SCH 23390. Brain Res. *375*, 291–301.

Schilman, E.A., Uylings, H.B.M., Graaf, Y.G. d, Joel, D., and Groenewegen, H.J. (2008). The orbital cortex in rats topographically projects to central parts of the caudate-putamen complex. Neurosci. Lett. *432*, 40–45.

Schoenbaum, G., Chiba, a a, and Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nat. Neurosci. *1*, 155–159.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. J. Neurophysiol. *80*, 1–27.

Shen, W., Flajolet, M., Greengard, P., and Surmeier, D.J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. Science *321*, 848–851.

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 Neurosci. *30*, 228–235.

Tecuapetla, F., Matias, S., Dugue, G.P., Mainen, Z.F., and Costa, R.M. (2014). Balanced activity in basal ganglia projection pathways is critical for contraversive movements. Nat. Commun. *5*, 4315.

Thorndike, E.L. (1898). Animal intelligence: an experimental study of the associative processes in animals (New York: New York : Macmillan).

Tolman, E.C. (1948). Cognitive maps in rats and men. Psychol. Rev. *55*, 189–208.

Vermeire, S., Audenaert, K., De Meester, R., Vandermeulen, E., Waelbers, T., De Spiegeleer, B., Eersels, J., Dobbeleir, A., and Peremans, K. (2012). Serotonin 2A receptor, serotonin transporter and dopamine transporter alterations in dogs with compulsive behaviour as a promising model for human obsessive-compulsive disorder. Psychiatry Res. - Neuroimaging *201*, 78–87.

Vertes, R.P. (2004). Differential Projections of the Infralimbic and Prelimbic Cortex in the Rat. Synapse *51*, 32–58.

Voorn, P., Vanderschuren, L.J.M.., Groenewegen, H.J., Robbins, T.W., and Pennartz, C.M.. (2004). Putting a spin on the dorsal–ventral divide of the striatum. Trends Neurosci. *27*, 468–474.

Wall, N.R., De La Parra, M., Callaway, E.M., and Kreitzer, A.C. (2013). Differential innervation of direct- and indirect-pathway striatal projection neurons. Neuron *79*, 347–360.

Washburn, M.S., and Moises, H.C. (1992). Electrophysiological and Morphological Amygdaloid Neurons in vitro Properties of Rat Basolateral. October *12*, 714–723.

Weiskrantz, L. (1956). Behavioral Changes associated with ablation of the amygdaloid complex in monkeys. J. Comp. Physiol. Psychol. *49 (4)*, 381–391.

Wilson, C.J. (1993). The generation of natural firing patterns in neostriatal neurons. Prog. Brain Res. *99*, 277–297.

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. Eur. J. Neurosci. *19*, 181–189.

Yin, H.H., Ostlund, S.B., Knowlton, B.J., and Balleine, B.W. (2005a). The role of the

dorsomedial striatum in instrumental conditioning. Eur. J. Neurosci. *22*, 513–523.

Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2005b). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur. J. Neurosci. *22*, 505–512.

Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav. Brain Res. *166*, 189–196.

Yin, H.H., Mulcare, S.P., Hilario, M.R.F., Clouse, E., Holloway, T., Davis, M.I., Hansson, A.C., Lovinger, D.M., and Costa, R.M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nat. Neurosci. *12*, 333–341.

Yu, C., Gupta, J., Chen, J.-F., and Yin, H.H. (2009). Genetic deletion of A2A adenosine receptors in the striatum selectively impairs habit formation. J. Neurosci. *29*, 15100–15103.

2

THE DOPAMINE TRANSPORTER GATES THE SWITCH BETWEEN GOAL-DIRECTED AND HABITUAL ACTIONS

"Sow a thought, and you reap an act;
Sow an act, and you reap a habit"

Samuel Smiles

SUMMARY

In an ever-changing world, animals are constantly learning new actions that lead to specific outcomes. With repetition, they can also form habits. Animals perform goal-directed actions and habits depending on the situation, highlighting the importance of balancing these two behavioural strategies. Although the circuits and mechanisms underlying learning and expression of goal-directed and habitual actions have been the subject of much scrutiny, less is known about how the brain balances the two strategies. It has been suggested that dopamine (DA) is essential for this process. Of particular interest is the dopamine transporter (DAT), which is critical for DA re-uptake and displays a gradient of expression in the striatum, with the highest expression in the striatal region implicated in habit formation (dorsolateral striatum - DLS). We investigated the role of DAT by training heterozygous DAT knock out ($DAT^{+/-}$) mice in a task in which they are trained on different lever-pressing schedules in different contexts, so that they are goal-directed in one and habitual in the other. We observed that while wild-type littermate ($DAT^{+/+}$) animals displayed goal-directed and habitual actions depending on the context, $DAT^{+/-}$ mice were habitual in both. Furthermore, $DAT^{+/-}$ animals were able to learn goal-directed actions when exposed to a single goal-directed inducing schedule. Therefore, these results suggest that the dopamine transporter is critical to balance the switch between goal-directed and habitual actions. These findings may have implications for understanding the circuitry underlying compulsive actions and persistent habits.

INTRODUCTION

Learning a new action by trial and error is crucial for animals to survive. With training and repetition, an action can become automatized and form a habit (Adams, 1982). Habits allow the brain to perform an action based on previous learned stimulus response (Yin and Knowlton, 2006), thus freeing different brain areas from the computational effort of a goal-directed tree search performance. We have all experienced goal-directed and habitual actions throughout life. However, an important part of this process is the capacity to switch between the two, depending on the situation. Previous studies have pinpointed the importance of dissociable striatal circuits in goal-directed actions versus habits; while animals with dorsolateral striatum (DLS) lesions show an increased sensitivity to devaluation of the outcome (behave as goal-directed) (Yin et al., 2004), lesions in the dorsomedial striatum (DMS) reduce sensitivity to the same devaluation paradigm (bias to habits) (Yin et al., 2005a, 2005b). However, the interaction between these circuits to balance the two behaviours is still largely unknown.

Dopamine has also been indicated as having a critical role in this phenomenon (Faure et al., 2005). However, the role of dopamine in different stages of learning in instrumental learning is complex, partially because there are many sources of variation in the control and regulation of (and by) dopamine. The major sources of striatal dopamine are the dopamine neurons from the midbrain, in particular from the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc) (Fallon and Moore, 1978). Dopamine projections to striatum follow a gradient, with DMS receiving a lateral VTA and ventromedial SNc projection, while DLS receives its dopamine inputs from the lateral part of SNc (Fallon and

Moore, 1978). In agreement with the projections, lesions of the nigrostriatal input to DLS have been shown to impair habit formation (Faure et al., 2005). Further, dopamine depleted mice are able to learn to correctly press an active lever upon viral injection that restores DA signal in either all of dorsal striatum, or only the DLS region (Darvas and Palmiter, 2009; Robinson et al., 2007). One theory put forward suggests dopamine as a good parent: it is needed during the initial learning stage, but once a strong stimulus-response association has been formed, it no longer requires dopamine to control actions (Horvitz et al., 2007).

While midbrain DA neurons have traditionally been associated with the transmission of an error prediction (PE) signal (Schultz, 1998), another type of response to eventful stimuli has been recorded in DA neurons, resembling more a motivational saliency signal (Matsumoto and Hikosaka, 2009). Interestingly, there is also a gradient of function within these 2 functions in the midbrain, with more motivational saliency neurons in dorsolateral SNc, and more PE related neurons in the ventromedial SNc and lateral VTA.

Of particular interest is the dopamine transporter (DAT), critical for dopamine re-uptake. The different dopamine neuronal projections create a gradient of DAT expression in dorsal striatum (Blanchard et al., 1994; Wickens et al., 2007; Shimada et al., 1992), with higher expression in DLS and lower in DMS (where COMT is very active (Matsumoto et al., 2003)). Amphetamine sensitization, a drug that interferes with DAT functioning (Kahlig et al., 2005), has been associated with faster transitions to habits (Nelson and Killcross, 2006). Furthermore, amphetamine exposure also leads to a decrease in spine density in DMS and an increase in spine density in DLS (Jedynak et al., 2007). DAT is, therefore, in a strategic position to be able to influence and control the dynamic changes between

DMS and DLS, and hence the balance between goal-directed actions and habits.

We investigated the role of DAT by training heterozygous DAT-KO ($DAT^{+/-}$) mice (Giros et al., 1996) and wild-type littermates ($DAT^{+/+}$) in the same behavioural task. We observed that while littermate control animals displayed goal-directed and habitual actions depending on the context, $DAT^{+/-}$ animals were habitual in both. These results do not seem to result from a lack of context discrimination nor from impairments in learning goal-directed actions, since $DAT^{+/-}$ were able to respond goal-directly during early training. Furthermore, when $DAT^{+/-}$ were trained in a single context in a schedule that traditionally renders wild-type animals' behaviour goal-directed, they were able to learn goal-directed actions. These results indicate that, in situations where there is competition between the networks, DAT is necessary to regulate the switch between the two behavioural strategies.

RESULTS

Both $DAT^{+/+}$ and $DAT^{+/-}$ animals acquire the lever pressing behaviour in the 2 schedules – 2 contexts task.

We trained mice to perform the same action (press a lever placed in the same location) for the same reinforcer, using either RI or RR schedules of reinforcement in two different contexts (**Figure 2.1a**). Mice were initially trained to lever press on a continuous reinforcement (CRF) schedule, with the potential to earn 5, 10 and 15 rewards per context, across 3 days. Then, mice underwent 2 days of RI30 (where the first press after an average 30 s interval had passed was reinforced) and RR10 (where on

average there was a reinforcement delivery every 10 presses) training, followed by 8 days of RI60 and RR20 training. Black6j mice (n=16) trained and tested in this 2 schedules – 2 contexts task showed a goal-directed control of the action in the RR context, but were habitual in the RI context (**Figure 2.1b,** RM two-way ANOVA, devaluation effect $F_{1,15}$=4.937, P=0.0421; schedule effect $F_{1,15}$=5.963, P=0.0275; interaction $F_{1,15}$=7.249, P=0.0167. One sample t-test against chance showed a difference of pressing only in the RR context – t=2.903, P=0.0109, but not in RI – t=0.9378, P=0.3632).

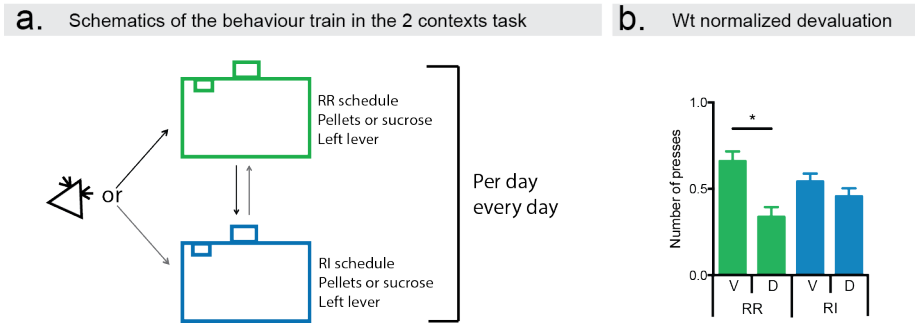

Figure 2.1 | The 2 schedules – 2 contexts task leads no animals that are goal-directed or habitual depending on the context. (a) Schematic of the two contexts task. Each animals was trained to perform an action in the same manipulandum (left lever) for the same reinforcer, but in different contexts that were associated with either RR or RI schedules **(b)** Normalized number of lever presses performed in the devaluation test (valued or devalued/sum of value and devalued) following 8 days of RR20/RI60 training, for wild-type, black6j animals (n=16). Results represent mean +/- SEM. *p<0.05

Next, we trained DAT heterozygote knock-out ($DAT^{+/-}$) animals and their respective littermates ($DAT^{+/+}$) in the same task. Both $DAT^{+/+}$ littermates (n=11) and $DAT^{+/-}$ (n=14) animals increased the pressing rate across days of training, in both contexts, and there were no significant differences in acquisition between experimental and control groups (**Figure 2.2a**, RR training: main effect of training $F_{13,299}$=25.94, P<0.0001; Group effect $F_{1,23}$

=0.3149, P=0.5801; Interaction $F_{13,299}$=0.4473, P=0.9509. **Figure 2.2b**, RI training: main effect of training $F_{13,299}$=28.02, P<0.0001; Group effect $F_{1,23}$ =0.2387, P=0.6297; Interaction $F_{13,299}$=0.4342, P=0.9566).

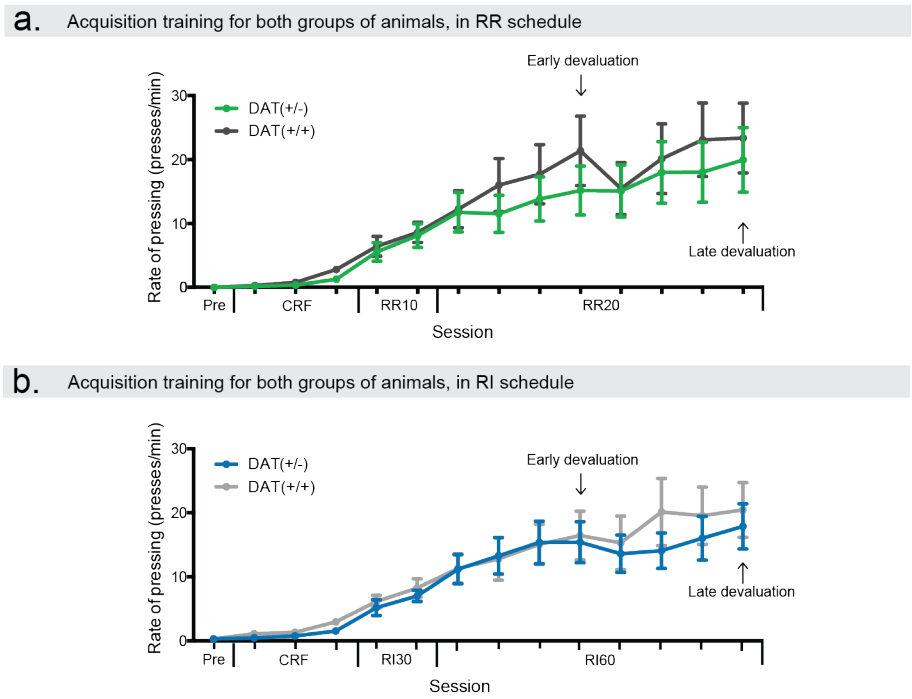

Figure 2.2 | DAT$^{+/-}$ and DAT$^{+/+}$ acquire lever pressing behaviour at the same rate. (a) Acquisition of the lever-pressing task in the controls (n=11) and the heterozygous deletion animals (n=14), in the RR-training context. **(b)** Acquisition of the lever-pressing task in the controls and the heterozygous deletion animals, in the RI-training context. Training days before satiety-specific devaluation tests are indicated. The rates presented were calculated dividing the number of presses by the length of the session, calculated per animal.

There was a difference between groups in the duration of the session across training (**Figure 2.3a**, main effect of training $F_{13,598}$=10.80, P<0.0001; group effect $F_{3,46}$=4.028, P=0.0126; interaction $F_{39,598}$=2.104, P=0.0001), but this was mainly due to differences early in the RR and RI training (posthocs not significant for the last days before both devaluations, and posthocs not significant between DAT$^{+/-}$ and DAT$^{+/+}$ in RR or RI).

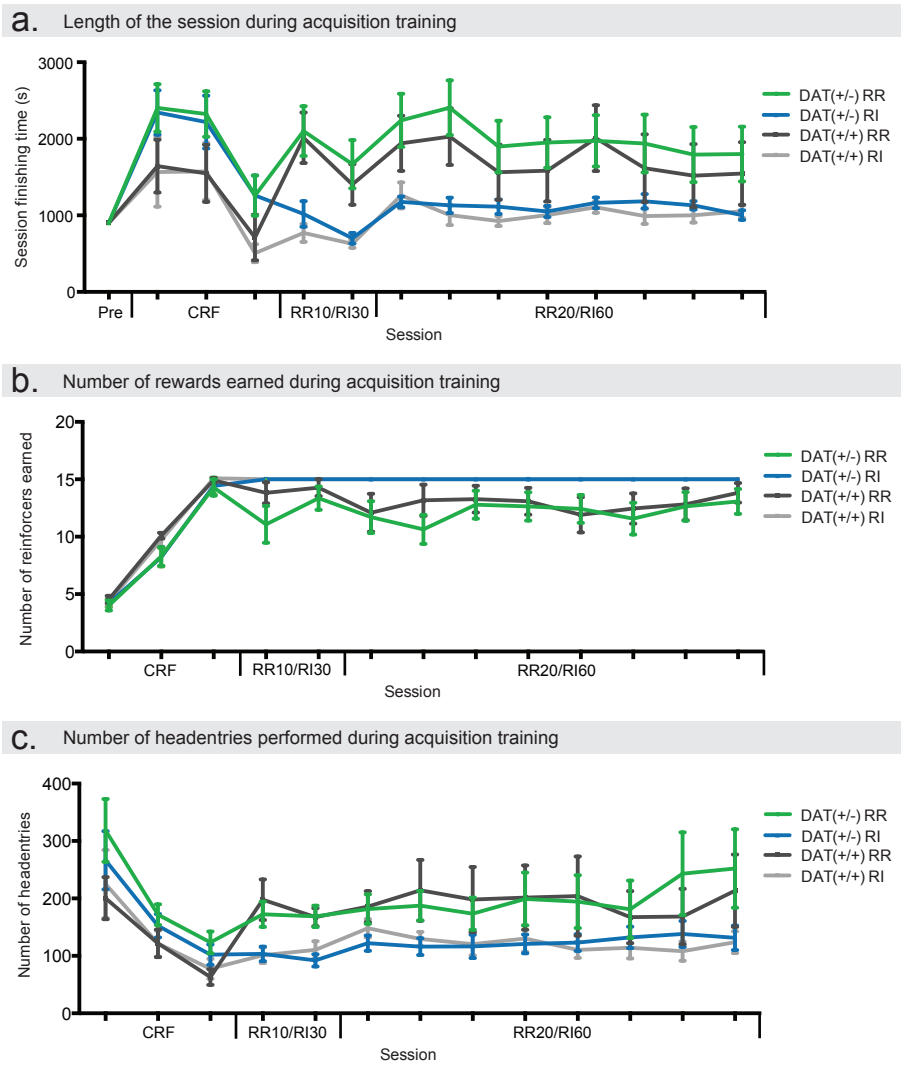

Figure 2.3 | $DAT^{+/-}$ and $DAT^{+/+}$ acquisition differ between contexts, but not experimental groups. (a) Length of each session for $DAT^{+/-}$ and $DAT^{+/+}$ in both contexts across training. **(b)** Number of headentries performed by $DAT^{+/-}$ and $DAT^{+/+}$ in both contexts across training. **(c)** Total number of earned reinforcers per context, for both $DAT^{+/-}$ and $DAT^{+/+}$ groups, across training.

There was also a significant difference between groups regarding the number of reinforcers earned per session (**Figure 2.3b**, main effect of training $F_{12,552}=90.41$, $P<0.0001$; group effect $F_{3,46}=3.572$, $P=0.0210$;

interaction $F_{36,552}$=1.807, P=0.0033), but this was again mainly due to differences early in RR and RI training (posthocs not significant for the last days before both devaluations, and posthocs not significant between $DAT^{+/-}$ and $DAT^{+/+}$ in RR or RI). Although we did not test differences in general locomotive effects in these animals, we observed that their number of headentries per session did not differ from control animals along training (**Figure 2.3c**, main effect of training $F_{12,552}$=64.92, P<0.0001; group effect $F_{3,46}$=2.470, P=0.0737; interaction $F_{36,552}$=0.9542, P=0.5476). We can conclude $DAT^{+/-}$ animals, at least at the macroscopic behavioural level, acquired the task with no significant different from the control littermates.

DAT-KO animals are impaired in goal-directed action performance

Since the behavioural mode controlling the animals' behaviour can't be retrieved by simple train analysis (as far as we can measure it), probing the behaviour is needed to elucidate it the action performed in each context is habitual or goal-directed. Thus, we used a devaluation test, as explained in the methods. This test allowed us to evaluate if the animals were behaving in a habitual or goal-directed manner on each context. We performed this test after 4 days of RR20/RI60 training (early devaluation) and 8 days of training (late devaluation). We normalized lever pressing number within session (**Figure 2.4b,d** – $Normalized_{Context\_A}$=(number of presses in context A in valued or devalued state)/(sum of presses in context A in both valuation states). In the early devaluation test, five $DAT^{+/-}$ and two five $DAT^{+/+}$ did not consume enough of the freely available reinforcer before testing, and hence had to be excluded from the analysis. As expected, $DAT^{+/+}$ control animals significantly changed their lever pressing behaviour in the RR, while maintaining their pressing in the RI context regardless of devaluation state (**Figure 2.4b**, RM two-way ANOVA, devaluation effect

$F_{1,16}$=15.05, P=0.0013; schedule effect $F_{1,16}$=0.9330, P=0.3485; interaction $F_{1,16}$=1.332, P=0.2654. Posthoc: RR, valued versus devalued – P<0.01, RI, valued versus devalued – ns). $DAT^{+/-}$ also displayed the same behavioural pattern, thus significantly changed their lever pressing behaviour in the RR, while maintaining their pressing in the RI context regardless of devaluation state (**Figure 2.4a**, RM two-way ANOVA, devaluation effect $F_{1,16}$=8.083, P=0.0118; schedule effect $F_{1,16}$=0.1634, P=0.6914; interaction $F_{1,16}$=0.4464, P=0.5136. Posthoc: RR, valued versus devalued – P<0.05, RI, valued versus devalued – ns). This result is in conformity with previous studies employing this 2 context-2 schedule task.

In the late devaluation test, we also observed that both $DAT^{+/-}$ and $DAT^{+/+}$ groups did not significantly decrease lever pressing in the RI context (**Figure 2.5a,b**); however, while the $DAT^{+/+}$ controls reduced lever pressing in the RR context following outcome devaluation, the $DAT^{+/-}$ animals pressed the same in the RR context, for both satiety states (**Figure 2.5b**, RM two-way ANOVA, devaluation effect $F_{1,20}$=12.56, P=0.0020; schedule effect $F_{1,20}$=0.3120, P=0.5827; interaction $F_{1,20}$=0.4284, P=0.5202. Posthoc: RR, valued versus devalued – P<0.05, RI, valued versus devalued – ns. **Figure 2.5a**: RM two-way ANOVA, devaluation effect $F_{1,20}$=3.963, P=0.0603; schedule effect $F_{1,20}$=0.1362, P=0.7159; interaction $F_{1,20}$=0.6008, P=0.4473. Posthoc: RR and RI, valued versus devalued – ns). Three $DAT^{+/-}$ had to be excluded from the analysis for lack of consumption of the freely available reinforcer before testing.

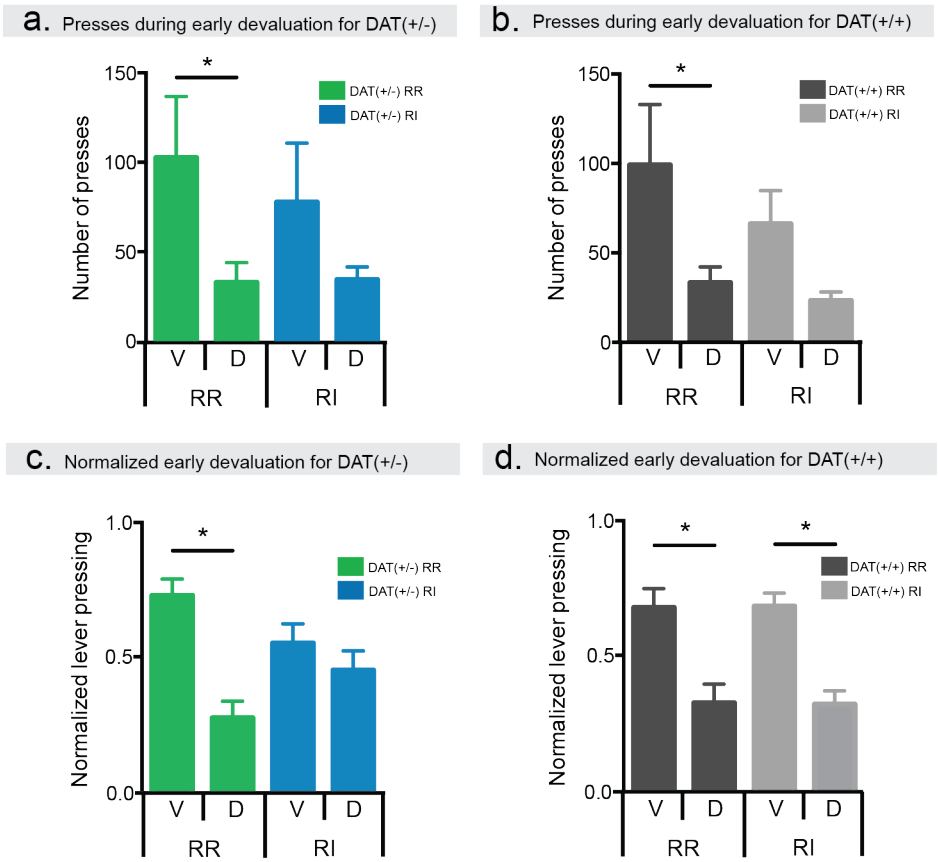

Figure 2.4 | Early in training both DAT$^{+/-}$ and DAT$^{+/+}$ are sensitive to outcome devaluation in the RR context, but not the RI. (a) Absolute number of lever presses performed in the devaluation test following 4 days of RR20/RI60 training. On the left panel are the results for the DAT$^{+/-}$ animals and on the right panel the DAT$^{+/+}$ animals. **(b)** Normalized number of lever presses performed in the devaluation test (valued or devalued/sum of value and devalues) following 4 days of RR20/RI60 training. On the left panel are the results for the DAT$^{+/-}$ animals and on the right panel the DAT$^{+/+}$ animals. Results represent mean +/- SEM. *p<0.05

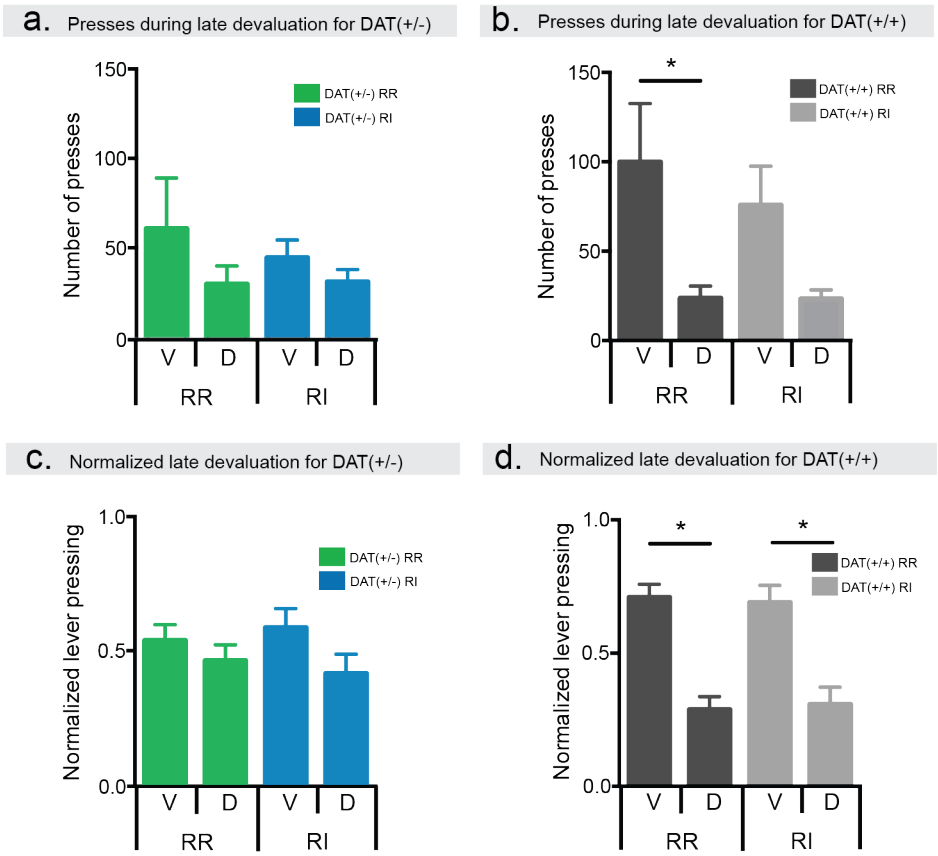

Figure 2.5 | Late in training DAT$^{+/-}$ is insensitive to changes in the value of the outcome in both training schedules. (a) Absolute number of lever presses performed in the devaluation test following 8 days of RR20/RI60 training. On the left panel are the results for the DAT$^{+/-}$ animals and on the right panel the DAT$^{+/+}$ animals. **(b)** Normalized number of lever presses performed in the devaluation test (valued or devalued/sum of value and devalues) following 8 days of RR20/RI60 training. On the left panel are the results for the DAT$^{+/-}$ animals and on the right panel the DAT$^{+/+}$ animals. Results represent mean +/- SEM. *p<0.05

DAT-KO animals are not more habitual in a single-context RR schedule

To further analyse the learning differences of DAT$^{+/-}$ animals, a new group of animals (16 DAT$^{+/-}$ and 14 DAT$^{+/+}$) was trained in only one schedule – RR – and one context.

Both groups acquired the lever pressing with no significant difference in lever rate (**Figure 2.6**, RM two-way ANOVA, Training effect $F_{21,588}$=23.54,

P<0.0001; group effect $F_{1,28}$=0.3286, P=0.5711; interaction $F_{21,588}$=1.007, P=0.4524). After the second devaluation test, both groups of animal decreased non significantly their lever pressing rate, possible due to the cumulative exposure to 4 days of extinction and 2 session of free access to the reinforcer they learned how to press for.

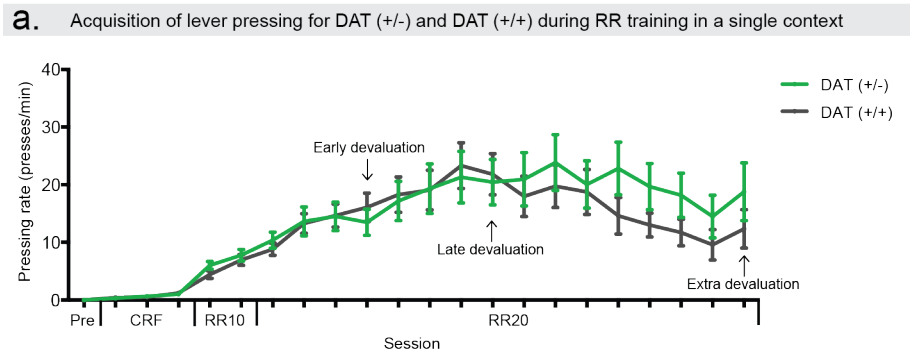

Figure 2.6 | DAT$^{+/-}$ and DAT$^{+/+}$ acquire lever pressing behaviour at the same rate in a single RR-training context. (a) Acquisition of the lever-pressing task in the controls (n=14) and the heterozygous deletion animals (n=16), under the RR schedule. Training days before satiety-specific devaluation tests are indicated. The rates presented were calculated dividing the number of presses by the length of the session, calculated per animal.

There was no difference between groups in the duration of the session across training (**Figure 2.7a**, main effect of training $F_{21,588}$=7.561, P<0.0001; group effect $F_{1,28}$=3.292x10^{-5}, P=0.9955; interaction $F_{21,588}$=1.472, P=0.0803). There was no significant difference between groups regarding the number of reinforcers earned per session (**Figure 2.7b**, main effect of training $F_{20,560}$=53.66, P<0.0001; group effect $F_{1,28}$=0.03812, P=0.8466; interaction $F_{20,560}$=0.8652, P=0.6325). Although we did not test differences in general locomotive effects in these animals, we observed that their number of number of headentries per session did not differ from control animals along training (**Figure 2.7c**, main effect of training $F_{20,560}$=3.762, P<0.0001; group effect $F_{1,28}$=0.5754, P=0.4545; interaction $F_{20,560}$=0.2617, P=0.9996).

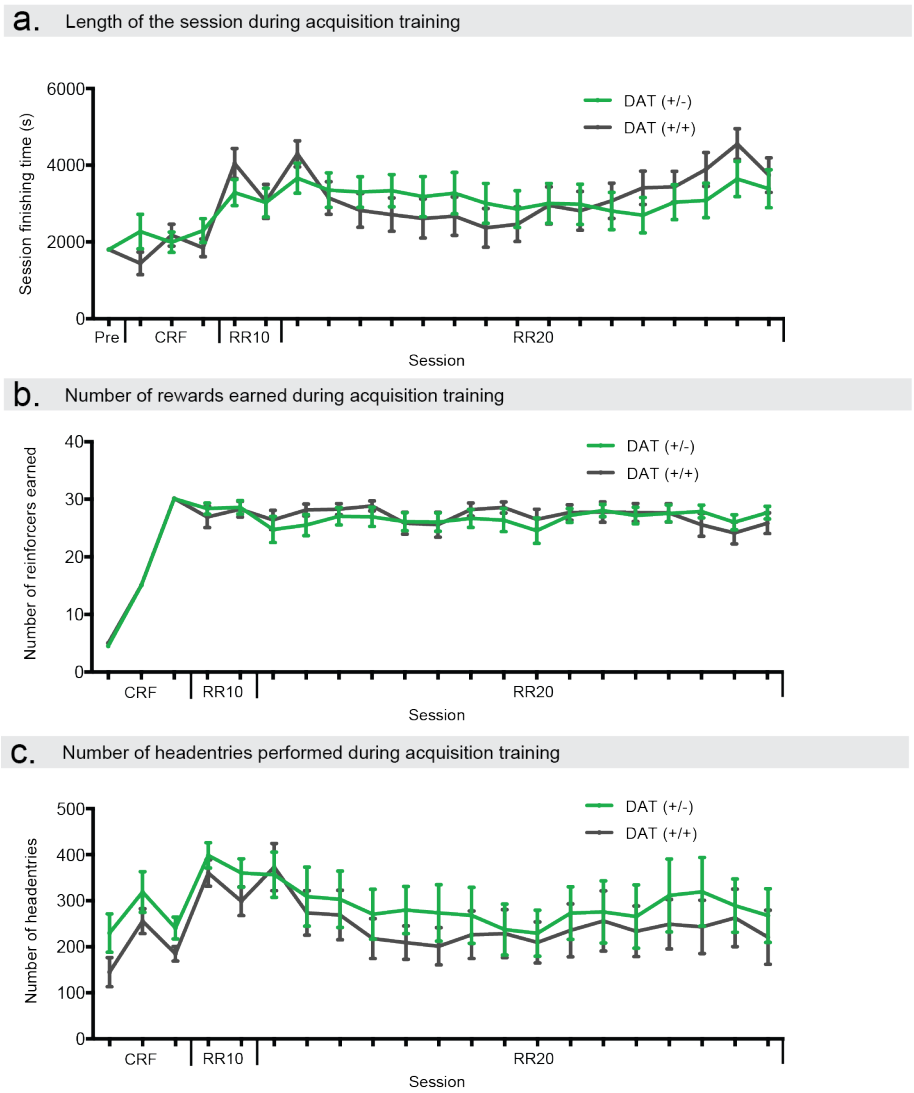

Figure 2.7 | DAT$^{+/-}$ and DAT$^{+/+}$ acquisition does not differ between experimental groups. **(a)** Length of each session for DAT$^{+/-}$ and DAT$^{+/+}$ in across training. **(b)** Number of headentries performed by DAT$^{+/-}$ and DAT$^{+/+}$ across training. **(c)** Total number of earned reinforcers, for both DAT$^{+/-}$ and DAT$^{+/+}$ groups, across training.

We observed that, as expected, DAT$^{+/+}$ controls were sensitive to manipulations in the expected value of the outcome, both after 4 and 8 of RR20 training (**Figure 2.8b**, RM 2-way ANOVA, Training effect $F_{2,26}$=0.6294, P=0.5408; devaluation effect $F_{1,13}$=20.09, P=0.0006;

interaction $F_{2,26}$=0.6136, P=0.5490. Posthoc: val vs deval, early <0.05, val vs deval, late <0.05). However, and unlike it was expected from the previous result, DAT$^{+/-}$ animals were also sensitive to this manipulation (**Figure 2.8a** RM 2-way ANOVA, Training effect $F_{2,30}$=0.6795, P=0.5145; devaluation effect $F_{1,15}$=13.64, P=0.0022; interaction $F_{2,30}$=1.054, P=0.3611. Posthoc: val vs deval, early <0.01, val vs deval, late <0.001).

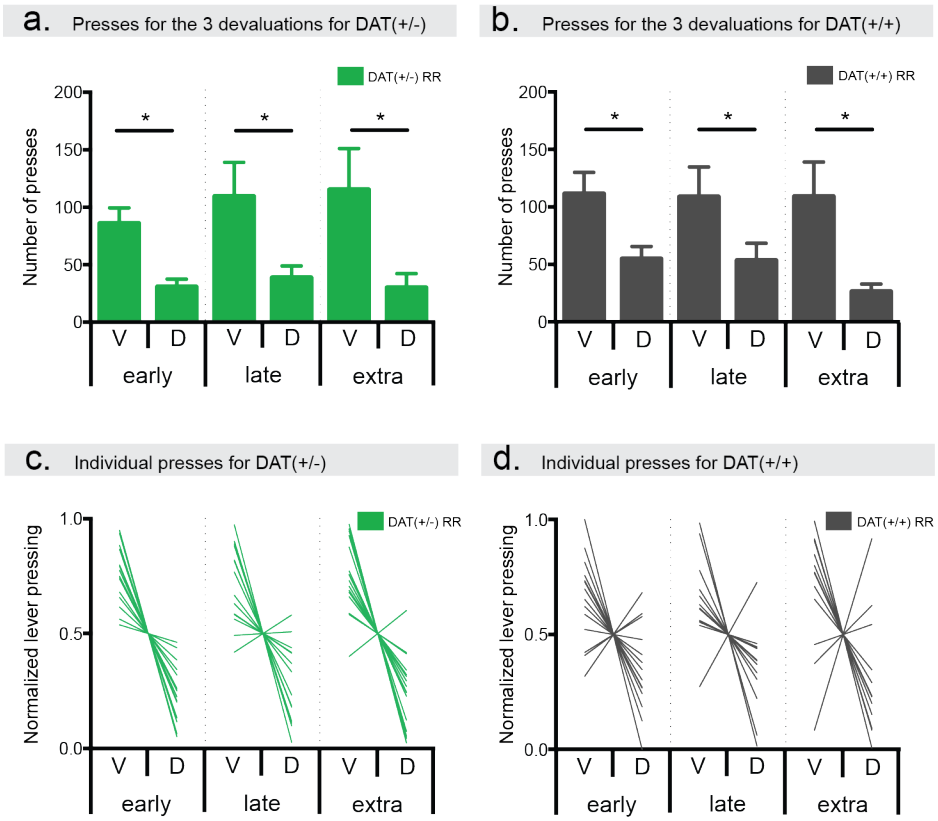

Figure 2.8 | DAT$^{+/-}$ are sensitive to changes in the value of the outcome for RR schedule, even after overtraining. (a,b) Absolute number of lever presses performed in the devaluation test following 4, 8 and 16 days of RR20 training for the (a) DAT$^{+/-}$ animals and the (b) DAT$^{+/+}$ animals. **(c,d)** Normalized number of lever presses performed in the devaluation test (valued or devalued/sum of value and devalues) following 4, 8 and 16 days of RR20 training. Individual traces represent each of the animals trained, for (c) DAT$^{+/-}$ animals and (d) DAT$^{+/+}$ animals. Results represent mean +/- SEM. *p<0.05

To further confirm this result, we trained both groups for an extra 8 days, with no change in the outcome of the devaluation test ($DAT^{+/+}$ posthoc val versus deval <0.001, $DAT^{+/-}$ posthoc val versus deval <0.0001).

Taken together, these results point to a role of DAT in gating these 2 behavioural strategies, rather than a simple role in habit formation regulation.

DISCUSSION

In this study we set of to investigate the role of the dopamine transporter in the balance between goal-directed and habitual actions. For that, we took advantage of a double context task developed previously in the laboratory (Gremel and Costa, 2013). To examine the role of this transporter we compared wild-type littermates ($DAT^{+/+}$) animals with mice that have a DAT heterozygous deletion (Giros et al., 1996) ($DAT^{+/-}$). These animals have shown to present a decreased rate of DA reuptake of about 50%, and have almost double the DA striatal concentration. We observed that after 8 days of train in RR20 and RI60, the $DAT^{+/+}$ control animals were sensitive to changes in the value of the outcome in the RR context, but insensitive in the RI context. On the other hand, the knock-out $DAT^{+/-}$ animals were habitual in both contexts, regardless of the associated training schedule. It would be possible that the result obtained in the devaluation test after 8 days of training is due to a lack of discrimination between both contexts in the $DAT^{+/-}$ group. However, an early devaluation test performed after only 4 days of RR20 and RI60 training showed that, at an initial training stage, both groups are goal-directed in the RR-associated context, and habitual in the RI-associated one.

Since DAT is one of the main intervenient in the cocaine and amphetamine drug sensitization (Kahlig et al., 2005; Ritz et al., 1987), and such sensitization bias animals towards habitual behaviour (Nelson and Killcross, 2006), the observation that $DAT^{+/-}$ animals are more habitual is expected. There are, however, at least two possible phenomenological explanations for this result. One possibility is that animals with the DAT heterozygous deletion progress to habitual states faster than wild-type animals. Yet, there is the possibility that these animals are not simply more habitual, or transit fast to a stimulus-response mode of action, but that there is a failing on gating between the two behavioural strategies once they are engaged in a habit. Since they were being trained under RI either right before or right after the RR training, maybe normal DAT functioning (and thus normal DA clearance) is needed to switch between the two similar manipulanda in different contexts. To access this difference, we trained a new set of animals in a 1 context - 1 schedule task, and observed that both the experimental group and the control littermates were sensitive to changes in the value of the outcome, even after extended 16 days of train (an extra 8 days compared to the previous task). Therefore, the result seen before is more likely to derive from the fact that each animal was trained sequentially in both schedules, rather than a faster transition to a habitual action. Since both tasks (one context versus two context) depend upon the same striatal circuitry (Gremel and Costa, 2013; Hilario et al., 2012; Yin and Knowlton, 2006), it is unlikely that the differences observed between experiments are due to differences in the circuitry responsible for the action learning in each case. A more parsimonious explanation is that during learning on both schedules, the exposure to the RI schedule and consequent changes in DA release and re-uptake is changing learning in the RR schedule.

One possibility is that RI training leads to differences in DA release and concentration within DLS. Because DA is rapidly cleared from the extracellular space in DLS, and since DA signalling is necessary for habit formation (Faure et al., 2005), it could justify why a habit takes time and repetition to form. It is possible that RI training leads to a specific pattern of DA release, and that the changes in the DA reuptake observed in the $DAT^{+/-}$ animals lead to a concentration of DA in DLS that would normally not be observed under RR training. Concentration changes could then be responsible for potentiating the circuit in DLS associated with performance of the RR lever press. Interestingly, uncertainty appears to be an important factor in the establishment of a habit (Derusso et al., 2010), and uncertainty variations can lead to different patterns of DA release (Fiorillo et al., 2003). Also consistent with these studies, it has been observed that during variable interval training there is a sharp increase of DA concentration in DLS, which is not observed in DMS (Shnitko and Robinson, 2015).

Although in both tasks animals had the opportunity to earn a total of 30 reinforcers, in the 2 contexts task there was a change of context each day, and only 15 reinforcers could be earned per context. Therefore, one can't exclude the hypothesis that the differences observed in sensitivity to devaluation during RR trained can be due to exposure to 2 different contexts, rather than training under 2 different schedules. It is also note notice that although below significant levels, the wild-type littermates are in general more goal-directed than expected. In fact, when the number of presses is normalized within animal/schedule, there are significant changes in the lever pressing behaviour in the devalued state, even in RI. We do not know why this is the case.

Deletions in DAT have been associated with spontaneous locomotor activity (Giros et al., 1996). However, this phenotype was observed in the homozygous mutant, but not in the heterozygous one. Nevertheless, and as a proxy of activity, we measured the length of the sessions and the number of headentries between experimental groups, and there was no significant difference between $DAT^{+/+}$ and $DAT^{+/-}$. Thus, the difference in pressing behaviour during different valuation states is probably not associated with a state of hyperactivity. This possibility can't, however, be completed excluded with this set of experiments. Noteworthy, the knockout animals used in the studies are constitutive deletions. Therefore, the results observed can also be due to changes in development, rather than changes in the circuit dynamics in adult mice.

Taken together, these findings further deepen our knowledge about the role of DA signalling in habit formation, and point to a role of DAT in the regulation of the transition between goal-directed and habitual states.

MATERIALS AND METHODS

Animals

All procedures were reviewed and performed in accordance with the Champalimaud Centre for the Unknown Ethics Committee guidelines, and approved by the Portuguese Veterinary General Board (Direcção Geral de Veterinária, approval 0421/000/000/2014). Male mice between 2 and 5 months of age were used in this study. DAT-KO mutants were generated as previously described (Giros et al., 1996). DAT-KO animals were obtained backcrossing mutant animals into C57Bl6/J background, and were bred with C57Bl6/J to obtain DAT-KO animals ($DAT^{+/-}$) and their littermate

controls ($DAT^{+/+}$). Animals were kept in a heterozygote state. Mice were housed between 2 and 4 animals per cage, under a 12 hours light/dark cycle. Experiments were performed on the light cycle. 11 $DAT^{+/+}$ and 14 $DAT^{+/-}$ were trained on the 2 schedules – 2 contexts task, and 14 $DAT^{+/+}$ and 16 $DAT^{+/-}$ were trained on the 1 schedule task.

Behavioural procedures

Training took place in behavioural chambers (Med-Associates, dimensions 23 cm x 20 cm x 12.7 cm – W x D x H) placed in sound attenuating boxes. Each chamber was equipped with a food magazine, a house light place on the wall on the left of the magazine and two retractable levers, one on each side of the magazine. MED-PC IV software was used to control the equipment and record lever presses, licks and head entries to the magazine.

2 schedules - 2 contexts training: each training session commenced with an illumination of the house light and left lever extension, and ended following schedule completion (15 reinforcers) or after 60 min, with the lever retracting and the house light turning off. The animals were trained to press the lever for an outcome of sucrose solution (20–30 µl of 20% solution per reinforcer) or sugar pellets (20 mg pellets, Bio-Serve formula F0071). The other reinforcer (not contingent on lever pressing) was provided in their home cage after training and used as a control for general satiation in the devaluation test. Before training started, mice were food restricted to 85% of their baseline weight, at which they were maintained for the duration of experimental procedures. Each day mice were trained in two separate operant chambers. For each mouse, the order of schedule exposure and the outcome obtained upon lever press were kept constant. However, mice were counterbalanced for starting context, schedule order

and reinforcer earned. On the first day, mice were trained to approach the food magazine (no lever present) in each context on a random time (RT) schedule, with a reinforcer delivered on average every 60 s for a total of 15 min. For the next 3 days, mice were trained in each context on a continuous reinforcement schedule (CRF), where every lever press made was reinforced, with an increasing number of possible earned reinforcers across training days (5, 10 and 15 per context; 10, 20 and 30 total each day). After CRF, mice were trained on random ratio (RR) and random interval (RI) schedules of reinforcement, with schedules differentiated by context, with the possibility of earning 15 reinforcers in each context or until 60 min had elapsed. Followed 2 days of RR10 (one reinforcer per average of 10 presses) and RI30 (one reinforcer available upon pressing, after a mean amount of 30 seconds has passed), training proceeded with 8 days of RR20 and RI60. Sensitivity to outcome devaluation was tested after 4 and 8 days of RR20/RI60 training.

Single context RR training: each training session commenced with an illumination of the house light and left lever extension, and ended following schedule completion (30 reinforcers) or after 90 min, with the lever retracting and the house light turning off. The animals were trained to press the lever for an outcome of sucrose solution (20–30 µl of 20% solution per reinforcer) or sugar pellets (20 mg pellets, Bio-Serve formula F0071). The other reinforcer (not contingent on lever pressing) was provided in their home cage after training and used as a control for general satiation in the devaluation test. Before training started, mice were food restricted to 85% of their baseline weight, at which they were maintained for the duration of experimental procedures. Each mouse was only trained in one context, keeping the reinforcer the same throughout training sessions. In the first day, mice were trained to approach the food magazine

(no lever present) in each context on a random time (RT) schedule, with a reinforcer delivered on average every 60 s for a total of 30 min. On the next 3 days, mice were trained on continuous reinforcement schedule (CRF), where every lever press made was reinforced, with an increasing number of possible earned reinforcers across training days (5, 15 and 30). After CRF, mice were trained on a random ratio (RR) schedule of reinforcement. Followed 2 days of RR10, training proceeded with RR20. Sensitivity to outcome devaluation was tested after 4, 8 and 16 days of RR20 training.

Outcome devaluation: outcome devaluation testing occurred across two consecutive days. In brief, on the valued day, mice had ad libitum access to the home cage outcome for 1 h before serial 5 minute, non-reinforced test sessions in the previous RI and RR training contexts (or in a single RR context for the single context animals). On the devalued day, mice were given 1 h of ad libitum access to the outcome previously earned by lever pressing, and then underwent serial non-reinforced test sessions in each training context. The order of context exposure during testing was the same as training exposure, with the order of revaluation day counterbalanced across mice.

Statistics

Statistical analyses were performed using GraphPad Prism 6 (GraphPad Software Inc., CA, USA). Repeated measures two way ANOVA were used to evaluate acquisition of lever pressing, headentries, total length of each session, number of earned reinforcers and lever pressing number within the devaluation test. These were followed by post hoc analyses when appropriate. Statistical significance was accepted for $p<0.05$. Results are presented as means±SEM.

ACKNOWLEDGEMENTS

We thank A. Vaz for colony management and G. Martins, C. Gremel and E. Dias-Ferreira for the valuable discussion about the protocol. This research was supported by the INDP Graduate Programme, a FCT fellowship to A.M.V. and an E.R.C. starting grant to R.M.C.

AUTHOR CONTRIBUTIONS

A.M.V. performed the behavioural experiments and analyses. A.M.V. and R.M.C designed the experiments.

REFERENCES

Adams, C.D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. Q. J. Exp. Psychol. Sect. B *34B*, 77–98.

Blanchard, V., Raisman-Vozari, R., Vyas, S., Michel, P.P., Javoy-Agid, F., Uhl, G., and Agid, Y. (1994). Differential expression of tyrosine hydroxylase and membrane dopamine transporter genes in subpopulations of dopaminergic neurons of the rat mesencephalon. Mol. Brain Res. *22*, 29–38.

Darvas, M., and Palmiter, R.D. (2009). Restriction of dopamine signaling to the dorsolateral striatum is sufficient for many cognitive behaviors. Proc. Natl. Acad. Sci. U. S. A. *106*, 14664–14669.

Derusso, A.L., Fan, D., Gupta, J., Shelest, O., Costa, R.M., and Yin, H.H. (2010). Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. Front. Integr. Neurosci. *4*, 1–8.

Fallon, J.H., and Moore, R.Y. (1978). Catecholamine innervation of the basal forebrain. IV. Topography of the dopamine projection to the basal forebrain and neostriatum. J. Comp. Neurol. *180*, 545–580.

Faure, A., Haberland, U., Conde, F., and El Massioui, N. (2005). Lesion to the nigrostriatal

dopamine system disrupts stimulus-response habit formation. J. Neurosci. *25*, 2771–2780.

Fiorillo, C.D., Tobler, P.N., and Schultz, W. (2003). Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. Science (80-.). *299*, 1898–1902.

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

Gremel, C.M., and Costa, R.M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nat. Commun. *4*, 2264.

Hilario, M.R.F., Holloway, T., Jin, X., and Costa, R.M. (2012). Different dorsal striatum circuits mediate action discrimination and action generalization. Eur. J. Neurosci. *35*, 1105–1114.

Horvitz, J.C., Choi, W.Y., Morvan, C., Eyny, Y., and Balsam, P.D. (2007). A "Good Parent" Function of Dopamine: Transient Modulation of Learning and Performance during Early Stages of Training. Ann N Y Acad Sci. *10*, 270–288.

Jedynak, J.P., Uslaner, J.M., Esteban, J. a., and Robinson, T.E. (2007). Methamphetamine-induced structural plasticity in the dorsal striatum. Eur. J. Neurosci. *25*, 847–853.

Kahlig, K.M., Binda, F., Khoshbouei, H., Blakely, R.D., McMahon, D.G., Javitch, J. a, and Galli, A. (2005). Amphetamine induces dopamine efflux through a dopamine transporter channel. Proc. Natl. Acad. Sci. U. S. A. *102*, 3495–3500.

Matsumoto, M., and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature *459*, 837–841.

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

Nelson, A., and Killcross, S. (2006). Amphetamine exposure enhances habit formation. J. Neurosci. *26*, 3805–3812.

Ritz, M.C., Lamb, R.J., Goldberg, S.R., and Kuhar, M.J. (1987). Cocaine receptors on dopamine transporters are related to self-administration of cocaine. Science (80-.). *237*, 1219–1223.

Robinson, S., Rainwater, A.J., Hnasko, T.S., and Palmiter, R.D. (2007). Viral restoration of dopamine signaling to the dorsal striatum restores instrumental conditioning to dopamine-deficient mice. Psychopharmacology (Berl). *191*, 567–578.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. J. Neurophysiol. *80*, 1–27.

Shimada, S., Kitayama, S., Walther, D., and Uhl, G. (1992). Dopamine transporter mRNA:

dense expression in ventral midbrain neurons. Mol. Brain Res. *13*, 359–362.

Shnitko, T.A., and Robinson, D.L. (2015). Regional variation in phasic dopamine release during alcohol and sucrose self-administration in rats. ACS Chem. Neurosci. *6*, 147–154.

Wickens, J.R., Budd, C.S., Hyland, B.I., and Arbuthnott, G.W. (2007). Striatal contributions to reward and decision making: Making sense of regional variations in a reiterated processing matrix. Ann. N. Y. Acad. Sci. *1104*, 192–212.

Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. *7*, 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. Eur. J. Neurosci. *19*, 181–189.

Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2005a). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur. J. Neurosci. *22*, 505–512.

Yin, H.H., Ostlund, S.B., Knowlton, B.J., and Balleine, B.W. (2005b). The role of the dorsomedial striatum in instrumental conditioning. Eur. J. Neurosci. *22*, 513–523.

3

AMYGDALA-STRIATAL INTERACTIONS IN INSTRUMENTAL LEARNING

"No man is an island,
Entire of itself,
Every man is a piece of the continent,
A part of the main."

John Donne

SUMMARY

In a changing environment, animals have to learn new actions constantly. A key component in learning new actions is reinforcement. In the case of model-free or habitual learning, dopamine has been implicated in this process. Dopamine signals the reward prediction error related to unpredicted rewards and unpredicted stimuli or actions that predict reward. This process seems less fit to modulate the learning of goal-directed actions, which are motivated by the expected value of the outcome. The orbitofrontal cortex and the basolateral amygdala (BLA), which project broadly to DMS but not DLS, have been postulated to play a role in signalling the expected value of outcomes resulting from goal-directed actions. In this study, we designed a set of experiments to investigate the role of BLA to DMS projections in learning novel goal-directed actions. We uncovered that stimulation of neurons projecting from BLA to DMS is sufficient to reinforce a novel, self-paced action. The action learned was highly sensitive to changes in the action-stimulation contingency, which suggests the animals were utilizing a goal-directed strategy. These findings, taken together with previous studies involving BLA in value tracking, suggest that the projections from BLA to DMS carry information necessary for the association between action and the current value of the outcome.

INTRODUCTION

The amygdala is a deep structure composed of several distinct nuclei that partake in the limbic system of the brain (LeDoux, 2007). Among these nuclei are the basolateral amygdala (BLA, which can be further subdivided into lateral, basal and basomedial amygdala) and the central nucleus (CeA, which can be further divided into lateral and medial central amygdala). Amygdala receives inputs from all sensory input modalities, and sends projections to several brain areas (Sah et al., 2003). The BLA – and its lateral subdivision in particular – is traditionally considered the sensory input of the amygdala. BLA projects broadly to striatum. Although these projections are broad, very interestingly they mostly avoid DLS, in particular the most antero-DLS (Kelley et al., 1982).

Amygdala has been described as important in a variety of behaviours related with emotional behaviour, such as fear, anxiety and stress, as well as in appetitive and aversive conditioning (Everitt et al., 2003; Janak and Tye, 2015; Johansen et al., 2011).

Studies have pinpointed that BLA is necessary for post-train devaluation of the US with LiCl in second order Pavlovian conditioning (Hatfield et al., 1996). Therefore, BLA appears to be involved in adjustment of responses to changes in the value of the US. In cats, physiological recordings suggest that there is an increase in coupling of gamma waves between BLA and striatum during learning (Popescu et al., 2009). Furthermore, disruption in BLA activity reduced gamma activity in striatum, and states of high striatal gamma correlated with increased coupling of BLA and striatal neuronal activity.

Recording from BLA and amygdala have shown that cells in this region develop a firing selectivity to stimuli associated with a specific reinforcer valence (Paton et al., 2006; Schoenbaum et al., 1998). Those signals were shown to be flexible, accompanied and correlated with learning, and are composed of separated neuronal populations that do not appear to be organized topologically

It was clear since earlier instrumental conditioning experiments – such as satiety specific devaluation – that when a new action-outcome is formed, the learned association is dependent on the expected value of the reinforcer (Dickinson, 1985). BLA-lesioned animals trained in a two actions – two outcomes task were able to acquire the instrumental task with no visible impairment, but were insensitive to devaluation and contingency degradation in a choice test (Balleine et al., 2003). Post-training BLA lesions also disrupt sensitivity to outcome devaluation (Ostlund and Balleine, 2008). Conversely, lesion in the anterior area of CeA led to animals that were sensitive to outcome devaluation in a training schedule that renders control sham animals habitual (Lingawi and Balleine, 2012). Functional disconnections between CeA and DLS, necessary for habitual action performance, also led to animals with increased sensitivity to outcome devaluation. Disconnections between BLA and DMS, as expected from previous results, led to animals that are insensitive to devaluation (Corbit et al., 2013). Taken together, this set of experiments exposes a possible dissociable and parallel network in amygdala, similar to that of dorsal striatum and cortex, responsible for different behavioural modes of instrumental learning. It also appears evident that amygdala is a key player in updating the value of an US, and associating changes of value in the US to associations with behaviour.

In this study, using a self-stimulation lever-pressing paradigm, we tested the sufficiency of BLA to DMS projection neurons to reinforce specific actions with the traits of a goal-directed strategy. Although previous retrograde labelling studies suggested that BLA neurons project mostly to D1 expressing medium spiny neurons within dorsal striatum (Wall et al., 2013), we observed that, in slice, both D1 and D2 expressing neuronal populations of DMS responded upon BLA stimulation. These results pave the way to our understanding of action value coding during goal-directed learning.

RESULTS

Stimulation of terminals of neurons projecting from amygdala to DMS is reinforcing.

To investigate the role of the neurons that project from BLA to DMS in action reinforcement, we trained animals in a self-reinforcing instrumental task. Black6-j mice with ChannelRhodopsin-2 (ChR, n=6) injections targeting BLA were trained to press a lever to receive a light stimulation in DMS (**Figure 3.1a**). As a control, Black6-j mice with YFP (n=6) injections were trained in the same manner. Since the spread of the injection extended past the BLA region, with this experiment we were only able to study the general projections from the amygdala area to DMS.

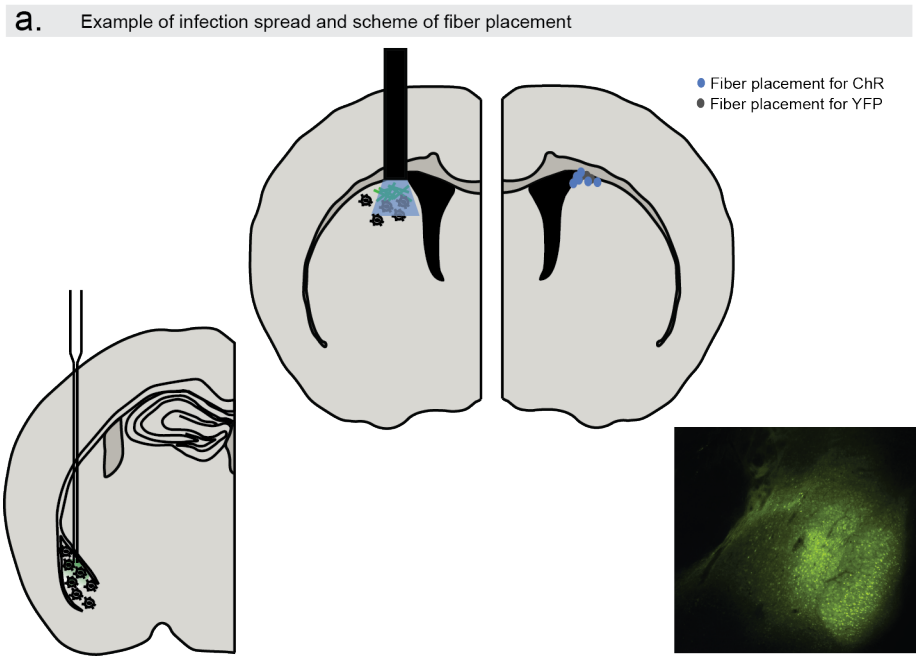

Figure 3.1 | **(a)** Schematics of the injection and fiber optics placement, representative histology slice of injection and fiber placement sites in DMS.

Two weeks after infection, animals were trained in an operant box with two levers: an active lever where pressing resulted in the delivery of blue light (473 nm) into DMS, and an inactive lever (no light delivered). Self-paced reinforced lever presses resulted in the delivery of 28 pulses of light (for 2 sec, at 14Hz, 10ms wide pulses). Each session lasted 60 minutes with no maximum number of reinforcers. In the session after each animal earned a cumulative number of reinforcers bigger than 400, they progress to one day of fixed ratio 2 (FR2), where every two presses led to the delivery of stimulation. If their pressing rate dropped on the FR2 day, the animal was kept on that schedule for another 3 days (1 out of 6 animals). If their pressing rate increased, they progressed to 3 days of fixed ratio 4 (FR4). All ChR-expressing mice increased the number of presses in the active lever with training, and pressed significantly more than YFP controls

(**Figure 3.2a right panel,** RM two-way ANOVA, training effect $F_{6,120}$=11.28, P<0.0001; group effect $F_{2,20}$=25.14, P<0.0001; interaction $F_{18,120}$=10.41, P<0.0001; posthoc of ChR active versus ChR inactive, YFP active and inactive P< 0.0001 after the 3rd day of train. **Left panel**, RM two-way ANOVA, training effect $F_{1,22}$=18.24, P=0.0003; group effect $F_{3,22}$=17.47, P<0.0001; interaction $F_{3,22}$=16.17, P<0.0001; posthoc of first versus last day of acquisition, ChR active - P< 0.0001, ChR inactive, YFP active and YFP inactive - ns). Different animals took a variable number of days to perform above 400 presses, taking a minimum of 3 days and a maximum of 6 (**Figure 3.2b**).

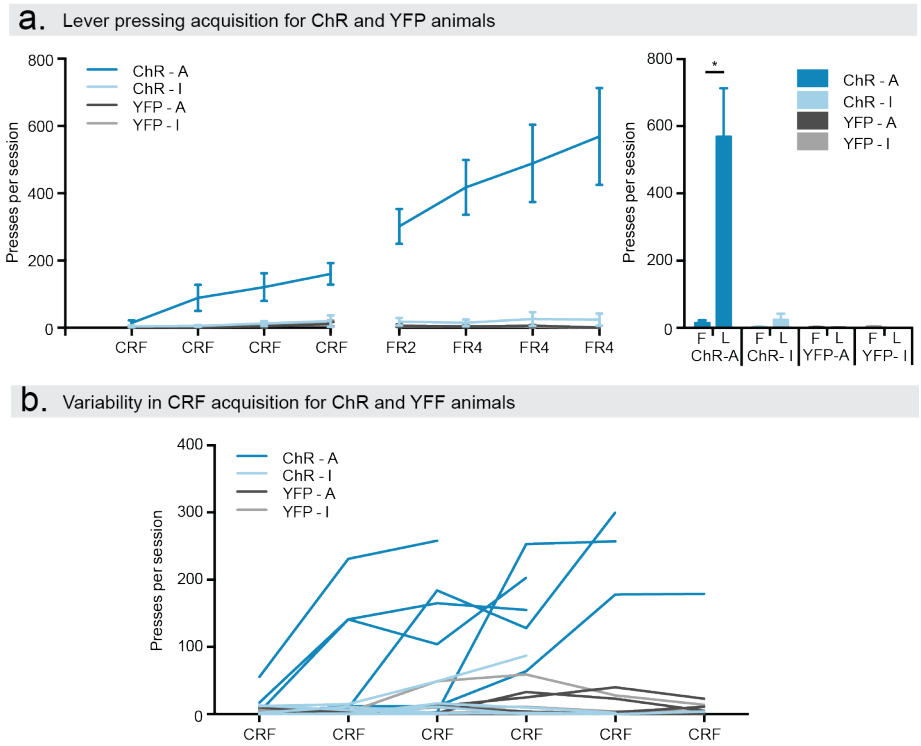

Figure 3.2 | Optogenetic self-stimulation of neurons projecting from amygdala to DMS supports the reinforcement of an action. **(a)** Acquisition of lever pressing in the active versus the inactive lever for ChR and YFP animals, throughout training (left panel) and in the first and last day of training (right panel). **(b)** First days of CRF acquisition of lever pressing in the active versus the inactive lever for ChR and YFP animals, until they reach are moved to the FR2 schedule. Results represent mean +/- SEM. *p<0.05

Self-stimulation of Amygdala-DMS projections is sensitive to contingency degradation.

To determine if the instrumental behaviour performed by the animals was goal-directed, we performed a contingency degradation test, to evaluate the sensitivity to changes of contingency between the action and the reinforcer. During contingency degradation (CD), the light stimulation was delivered non-contingently upon lever pressing, with the same probability of reinforcement per unit of time as during training. Following CD, animals underwent contingency reinstatement, where pressing the active lever would again lead to the delivery of stimulation. ChR infected animals decreased the number of presses during the CD session (**Figure 3.3a**, two way RM ANOVA, effect of CD $F_{2,44}$=10.75, P=0.0002; Group effect $F_{3,22}$=18.38, P<0.0001, interaction $F_{6,44}$=10.56, P<0.0001. Posthoc last day versus CD for active lever ChR animals - P<0.0001), and resumed their lever pressing behaviour during reinstatement (posthoc CD versus reinstatement for active lever ChR animals - P<0.0001, posthoc last day versus reinstatement for active lever ChR animals - ns). Individual results confirm the group results **(Figure 3.3b)**. These results suggest the animals are performing their behaviour in a goal-directed mode. Further analysis suggests that while in the first 10 minutes of degradation the animals appear to press at rates similar to the last day of training, it is followed by a sharp decrease of lever pressing behaviour that is sustained throughout the rest of the session (**Figure 3.3c**).

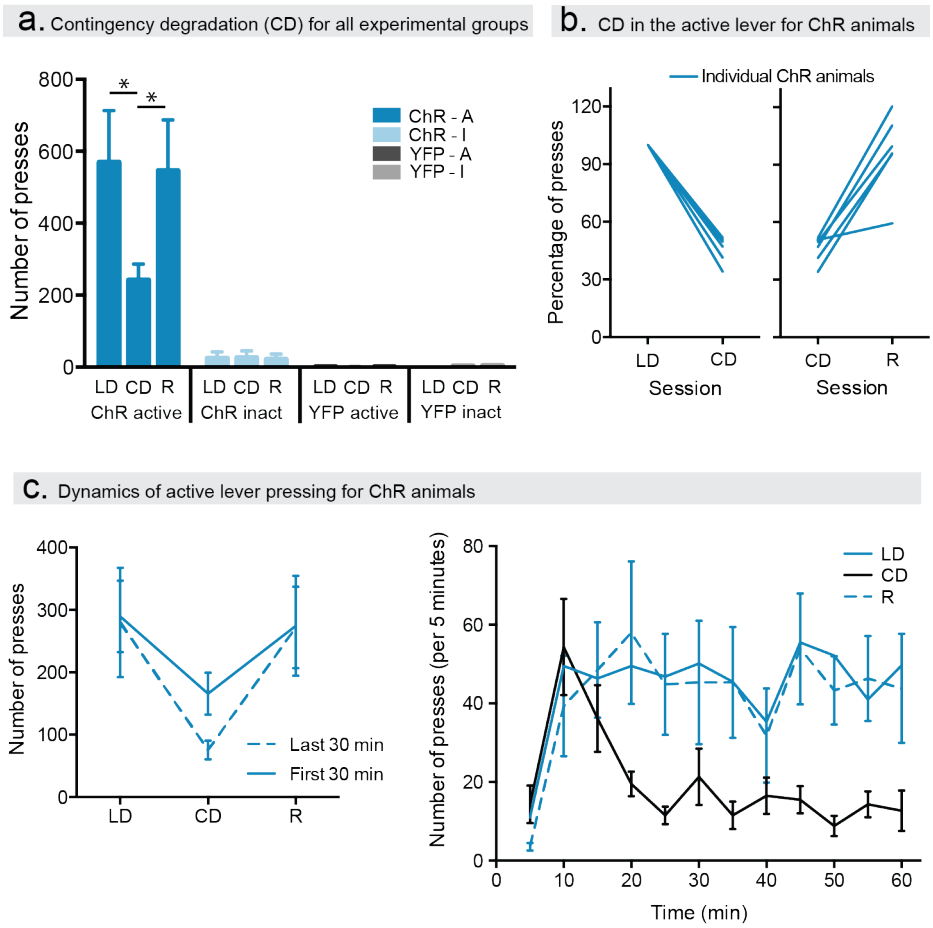

Figure 3.3 | Amygdala-DMS neuronal stimulation is sensitive to contingency degradation. **(a)** Total number of presses during the last day of training (LD), contingency degradation session (CD) and reinstatement session (R), for the experimental and the control groups. **(b)** Individual lever pressing percentage (from last day of training) for the experimental animals, in the active lever, for the CD (left panel) and the R (right panel) sessions. **(c)** Temporal dynamics in the active lever for ChR animals during the contingency degradation session (Left panel) and number of presses for the 3 sessions, separated by the first 30 and the last 30 minutes of the session (right panel). Results represent mean +/- SEM. *p<0.05

Animals self-stimulating amygdala-DMS projection neurons can learn reversals in action-stimulation contingency.

After the reinstatement we tested the dynamic capacity of animals to follow the reinforcer by swapping the contingency of both levers. In the

first day of switch, pressing the lever that was previously inactive led to delivery of stimulation, while pressing the previously active lever led to no outcome delivery. ChR infected animals decreased their lever pressing behaviour in the previously active lever within one session, while increasing the rate of pressing the previously inactive one (**Figure 3.4a,** two way RM ANOVA, switch effect $F_{3,15}$=3.728, P=0.0349; lever effect $F_{1,5}$=59.68, P=0.0006; interaction effect $F_{3,15}$=22.84, P<0.0001, posthoc FR4 versus sCRF, sFR2 and sFR4 for both lever - P<0.05).

By analysing the cumulative number of presses in both levers (**Figure 3.4b**), it is possible to observe that while presses in the previously active lever increased in the beginning of the session, by around 20 minutes pressing in both levers was equivalent and half way through the end of the session the presses in the previously inactive had already exceeded the presses in the previously active lever. The probability of pressing the previously active lever showed a sharp decrease over time within the first session of switch training (**Figure 3.4c**).

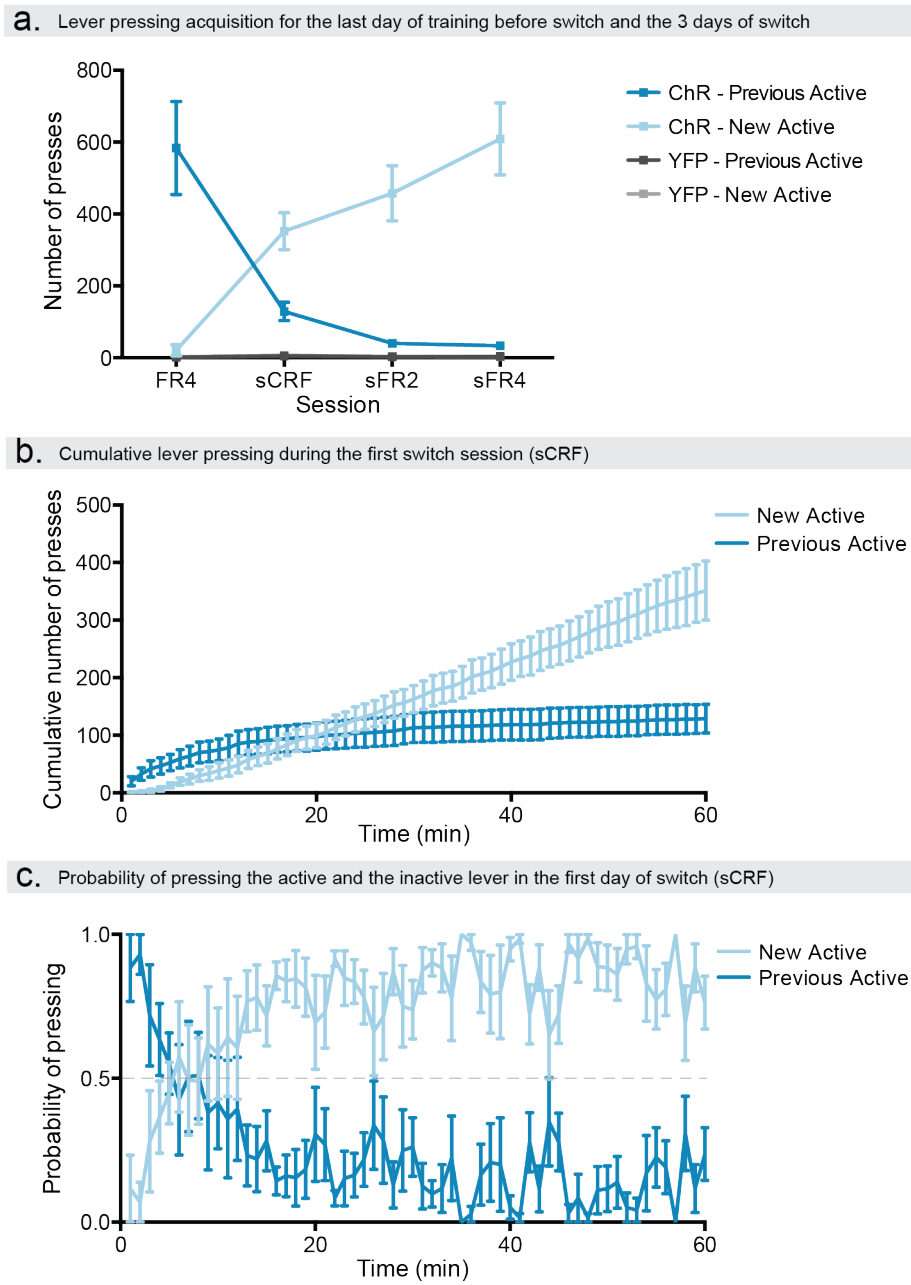

Figure 3.4 | ChR infected animals learn to switch pressing to the inactive lever upon changes in contingency between both levers. (a) Lever pressing behaviour for ChR and YFP infected animals during the last day of training before switch, and the next 3 switch days. **(b)** Cumulative number of lever presses during the first day of switch for ChR infected animals, in the previously active (now inactive) and the previously inactive (now active) levers. **(c)** Probability of pressing the previously active (now inactive) and the previously inactive (now active) levers the first day of switch for ChR infected animals.

Animals self-stimulating amygdala-DMS projection neurons can dynamically track action-stimulation contingency.

To further explore these dynamics of behaviour flexibility, after switch training the animals were trained in block sessions, where the active and the inactive levers would switch within session, dependent on the number of reinforcers earned by the animal. In an x-reinforcer block training, the active and the inactive lever changed after x reinforcers were earned. The number of reinforcers per block was 50 or 100, depending on the lever-pressing rate of the animals before the start of the block training, in order to ensure that they could perform at least 4 blocks per session. Animals were trained for a minimum of 3 days, and got an extra day if their percentage of corrected presses did not reach 70%. One animal was excluded from this analysis due to a program error on the first day of training, although he progressed to behave at the same performance rate as the rest of the group.

To observe in differences between the first and the last day of training in detail, one can look at an example animal (**Figure 3.5a,b**). This animal performed 6 complete blocks on both the first and the last day of training. However, while in the first blocks of day 1 he did not alternate between levers, simply pressing both, an alternation behaviour emerged by the end of day one. On the last day of block training the animal was alternating between levers, completely following the active lever changes.

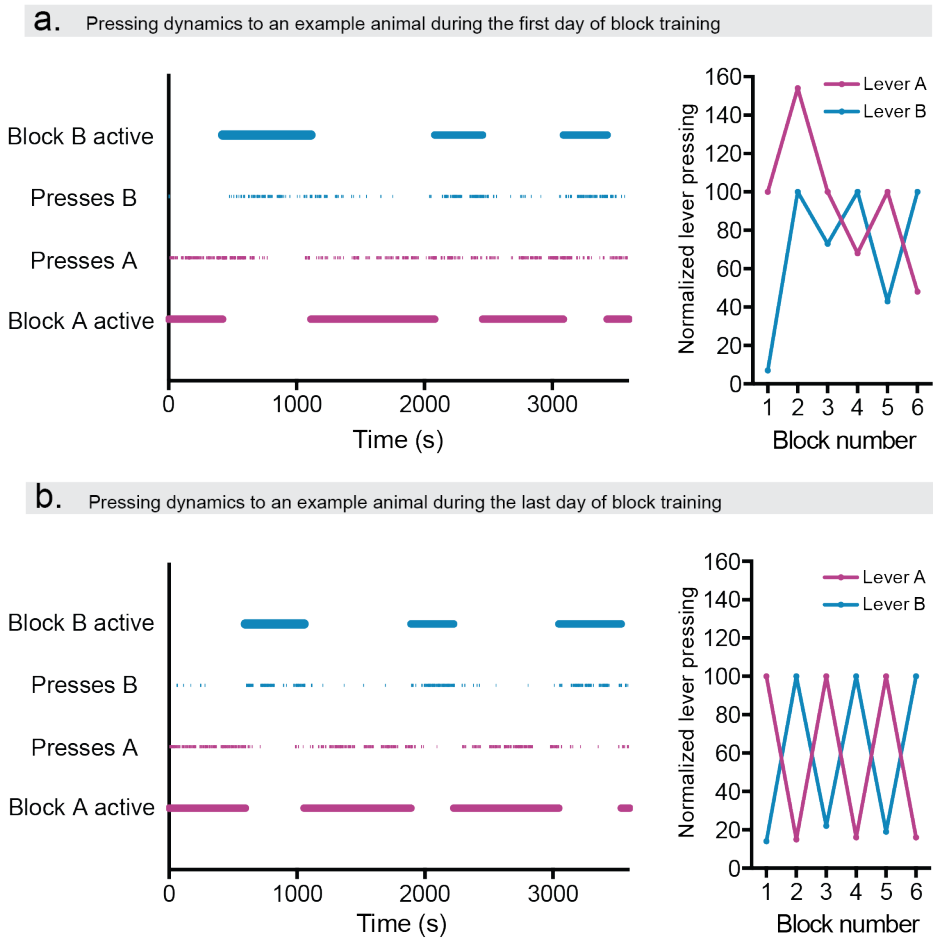

Figure 3.5 | Macrostructure of the behaviour of an example animals during block training. (a) Lever pressing structure of an example animal on the first day of block training (left panel). Number of presses on each completed block, for each lever, normalized by the number of active presses performed during that block (right panel). **(b)** Lever pressing structure of an example animal on the last day of block training (left panel). Number of presses on each completed block, for each lever, normalized by the number of active presses performed during that block (right panel). On the left panels, full line represent current active lever in the block, and ticks represent presses in either lever. Different colours represent different levers, regardless of active or inactive status within the block.

Different animals completed a different number of blocks per session. The maximum number of blocks completed was 9, and the minimum was 3. To compare all the animal together, we can compared the first 3 blocks of the first training day, and the first 3 blocks of the last training day. This is a

way of comparing average results for a number of blocks that every animal completed. As it is easily observed, animals in the first day of training did not alternate between the two levers, showed increases and decreases in lever pressing behaviour similar to both levers and there was no significant differences between blocks for either lever (**Figure 3.6a,** two-way RM ANOVA**,** Block factor $F_{2,8}$=2.558, P=0.1384; lever effect $F_{1,4}$=2.116, P=0.2195, interaction $F_{2,8}$=0.4209, P=0.6702). Furthermore, they appear to have a preference for pressing the active lever of the first block. If one looks at the last day of training, however, the alternation behaviour observed in the individual animal example appears to be common for all the animals, and they appear to be following the current active lever rather than pressing both continuously (**Figure 3.6b, left panel,** two-way RM ANOVA**,** block factor $F_{2,8}$=1.321, P=0.3194; lever effect $F_{1,4}$=37.80, P=0.0035, interaction $F_{2,8}$=27.02, P=0.0003; posthoc 1st and 3rd blocks versus 2nd – P<0.05 for both levers). When looking at all blocks completed by all animals in the last day of training, it is observed that this behaviour is consistent across the full training session (**Figure 3.6b**, **right panel**).

Another way of analysing the dynamics of lever pressing in both levers is to evaluate the probability of pressing one lever versus the other. Because different blocks take different times to complete (both between animals, and also within animal, between block number), to have a comparable measure of probability of pressing, the blocks were separated in 10 equal time intervals, and the probability of pressing was calculated for both lever, for each interval in each block. In both the first and the last day of training both probabilities fluctuated dependently on the contingency of the block. Nevertheless, in the last day there were sharper lever pressing changes upon block switch, and an overall more consistent difference

between the probabilities of pressing the current active lever versus the previous one (**Figure 3.6c**).

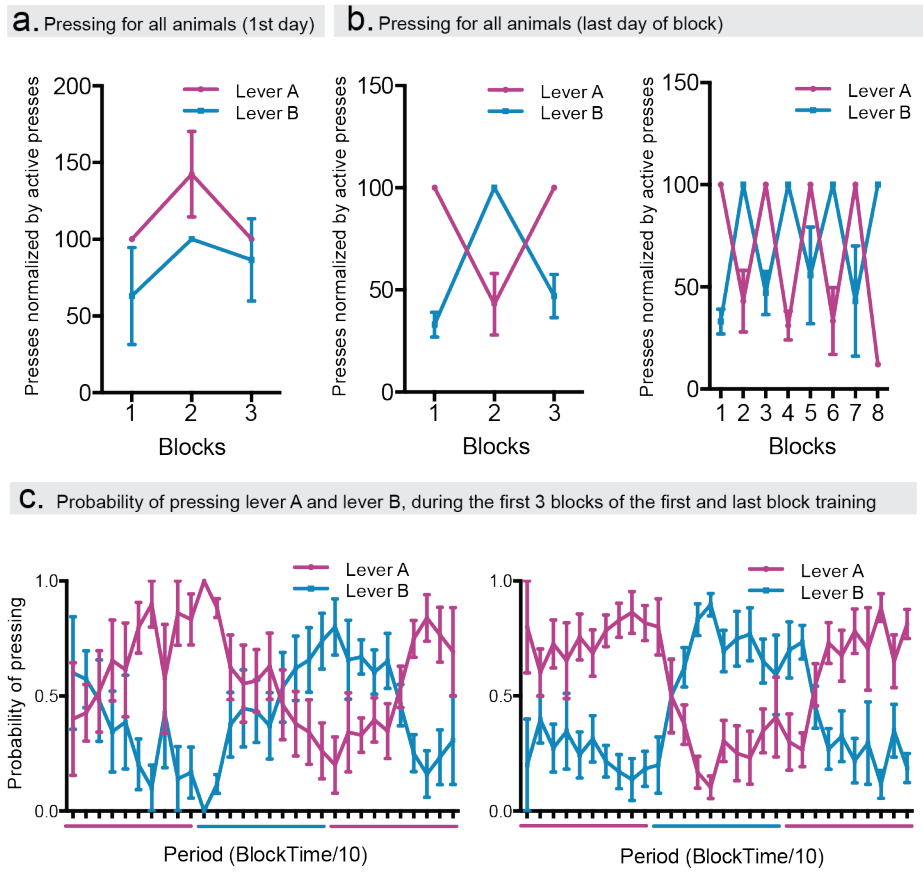

Figure 3.6 | Lever pressing behaviour for all ChR animals, during block training. (a) Number of presses on each completed block, for each lever and for the first 3 completed blocks of all animals, normalized by the number of active presses performed during that block, on the first day of block training. **(b)** Number of presses on each completed block, for each lever and for the first 3 completed blocks of all animals (left panel) or for all completed blocks (right panel), normalized by the number of active presses performed during that block, on the first day of block training. **(c)** Probability of pressing each lever, calculated in 10 bins within each block, for the first (left panel) and the last (right panel) day of training.

The number of blocks performed within each session did not significantly changed between the first and the last days of block training (**Figure 3.7a,** 2-tailed, paired t-test t=2.138, df=4, P=0.0993). There was, however, a

tendency to decrease the number of presses in the previously active lever upon block changing (**Figure 3.7c,** 2-tailed, paired t-test t=2.052, df=4, P=0.1095), and there was a significant decrease in both the time interval between block change and the first active press in the new block (**Figure 3.7b,** 2-tailed, paired t-test t=2.869, df=4, P=0.0455), and also in the total number of wrong presses per session (presses in the now inactive lever) (**Figure 3.7d,** 2-tailed, paired t-test t=2.775, df=9, P=0.0216). All this results suggest animals were capable of learning and performing block training.

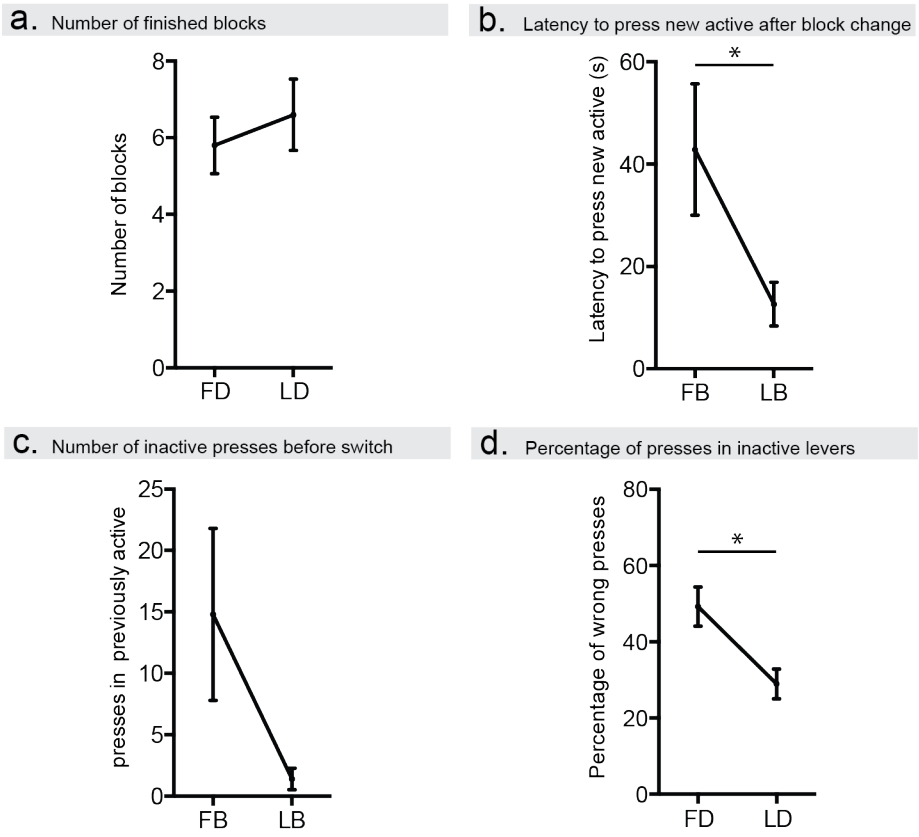

Figure 3.7 | Amygdala-DMS projection self-stimulating animals learn to perform block sessions, improving between the first and the last day. (a) Number of completed blocks. **(b)** Mean time (in seconds) taken for the animal to press the new active lever, upon block transition. **(c)** Average number of presses in the previously active (now inactive) lever, before a press in the new active one, upon block transition. **(d)** Percentage of rewarded (in

the current active) versus non-rewarded (presses in current inactive) presses. FD – first day of block training; LD – last day of block training. Results represent mean +/- SEM. *p<0.05

Stimulation of terminals of DMS projecting BLA excitatory neurons is reinforcing.

Since the expression of ChR around BLA was very pronounced in the WT injected animals, we decided to use a recently developed Cre line, NL189, which in amygdala expresses Cre recombinase preferentially in BLA (**Figure 3.8a**). Preliminary inspection with injections of a Cre-dependent tdTomato and a YFP under the control of a CamKII promoter showed that the cells expressing Cre-recombinase are likely glutamatergic neurons (**Figure 3.8b**).

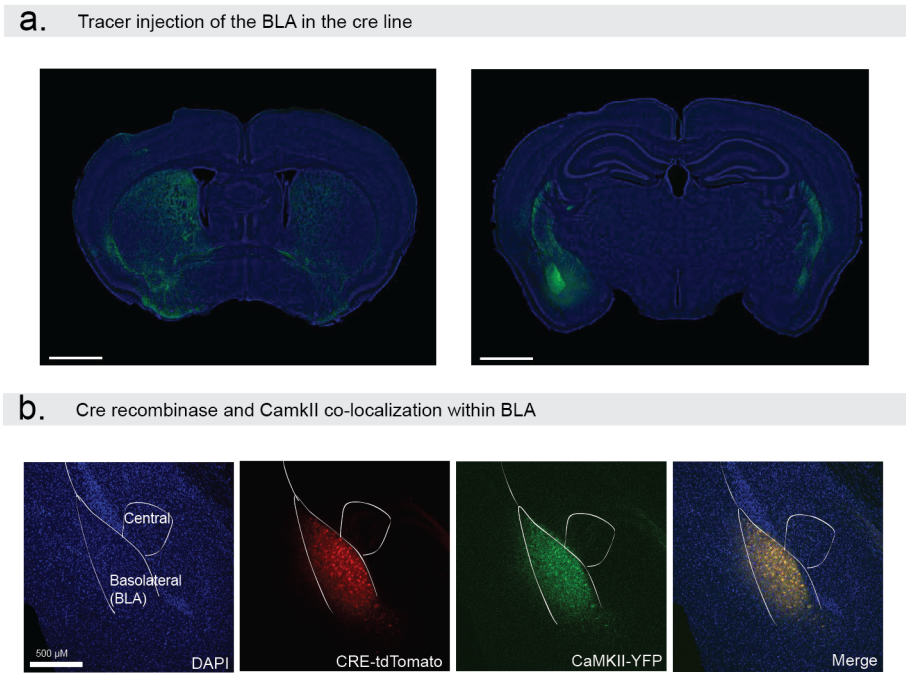

Figure 3.8 | **(a)** Expression of an AAV tracer in the BLA of a NL-189 Cre recombinase animal, and respective expression of fibers in DMS, from Biolucida. **(b)** Co-localization of expression of a Cre dependent tdTomato marker and a YFP marker under the control of the CamKII promoter.

The information about this line is available at the GENSAT Project, and available at the Mutant Mouse Resource and Research Centers, under the name Tg(Arhgef6-cre)NL189Gsat/Mmucd.

To investigate the role of the neurons that project from BLA to DMS in action reinforcement, we trained animals in a self-reinforcing instrumental task. NL189 mice with ChR under the control of a Cre recombinase (AAV2/9) viral injections targeting BLA were trained to press a lever to receive a light stimulation in DMS (**Figure 3.9a**). Control NL189 animals were injected with a virus with a Cre dependent expression of YFP. Two weeks after infection, animals were trained in an operant box with two levers: an active lever where pressing resulted in the delivery of blue light (473 nm) into DMS, and an inactive lever (no light delivered). Self-paced reinforced lever presses resulted in the delivery of 28 pulses of light (for 2 sec, at 14Hz, 10ms wide pulses). Each session lasted 60 minutes with no maximum number of reinforcers. All animals were trained for 11 days in a CRF schedule of reinforcement.

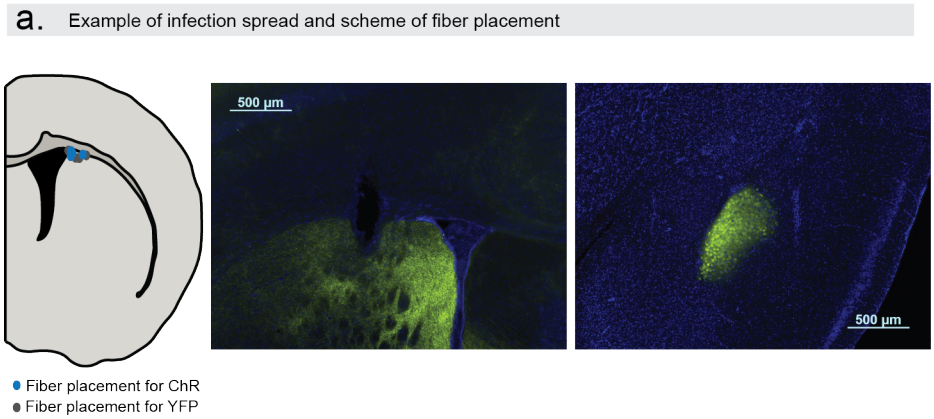

Figure 3.9 | **(a)** Scheme of fiber optics placement, and representative viral expression in projections in DMS and in BLA neurons.

ChR-expressing mice (n=5) increased the number of presses in the active lever with training, and pressed significantly more than YFP controls (n=4) (**Figure 3.10a,** two-way RM ANOVA, train effect $F_{10,140}$=2.563, P=0.0071; group effect $F_{3,14}$=2.252, P=0.1272; interaction effect $F_{30,140}$=2.844, P<0.0001; posthoc ChR active versus ChR inactive, YFP active and YFP inactive – P< 0.05 for the last 4 days of acquisition). ChR-expressing animals also press significantly more the active lever in the last day of training, comparing to the first (**Figure 3.10b,** two-way RM ANOVA, train effect $F_{1,14}$=3.098, P=0.1002; group effect $F_{3,14}$=3.425, P=0.0468; interaction effect $F_{3,14}$=3.188, P=0.0567; posthoc ChR active first versus last day – P<0.05).

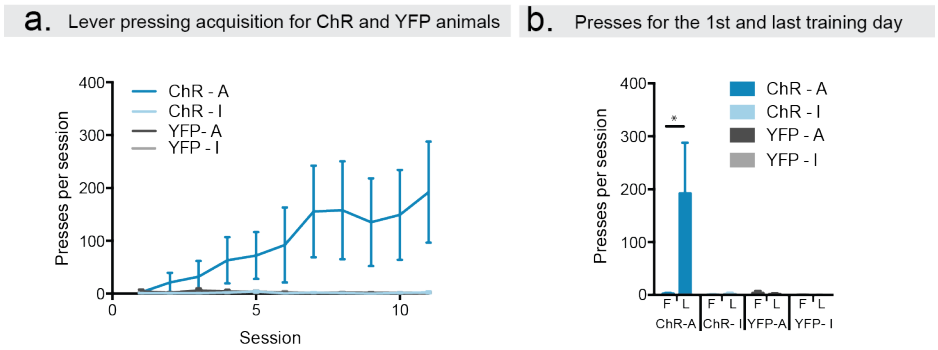

Figure 3.10 | Optogenetic self-stimulation of neurons projecting from BLA to DMS supports the reinforcement of an action. (a) Acquisition of lever pressing in the active versus the inactive lever for ChR and YFP infected animals, throughout training. **(b)** Number of lever presses during the first and last day of training for ChR and YFP infected animals. Results represent mean +/- SEM. *p<0.05

After the last day of acquisition, 2 ChR infected animals started having seizures and had to be excluded from the experiment. To evaluate if the instrumental behaviour performed by the animals was goal-directed, we performed a contingency degradation test, as described before. ChR infected animals decreased the number of presses during the CD session

(**Figure 3.11a**), and resumed their lever pressing behaviour during reinstatement, suggesting the animals are performing their behaviour in a goal-directed strategy. However, no conclusion can be taken out of these observations, since they result from a very reduced number of animals. Following reinstatement we trained animals in block training (**Figure 3.11b**). One animal stopped pressing after the reinstatement day, and was excluded from the following analysis. Therefore, only two animals underwent block training. As can be observed, after 4 days of block training the ChR infected animals learned to alternate between levers.

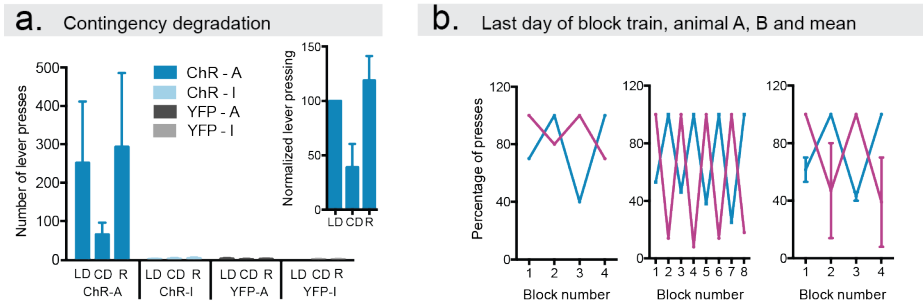

Figure 3.11 | Animals receiving stimulation in BLA to DMS projecting neurons can follow changes in contingency. (a) Total number of presses and normalized lever pressing during the last day of training (LD), contingency degradation session (CD) and reinstatement session (R), for the experimental and the control groups. **(b)** Percentage of presses in both levers for the 2 animals trained in block training and their average, on the last day of block train.

As previously mentioned, BLA projects broadly to DMS (Kelley et al., 1982), and it is thought that those projections target mostly D1-expressing medium spiny neurons (Wall et al., 2013). To further explore the interaction between these two regions involved in goal-directed action, we decided to record either D1 or D2 cells in slice, while stimulation BLA infected cells with ChR (**Figure 3.12a**). We observed that neurons in both

D1 and D2 neuronal populations responded to stimulation (5 out of 7 recorded D1 neurons; 6 out of 9 recorded D2 neurons) (**Figure 3.12b,c,d**).

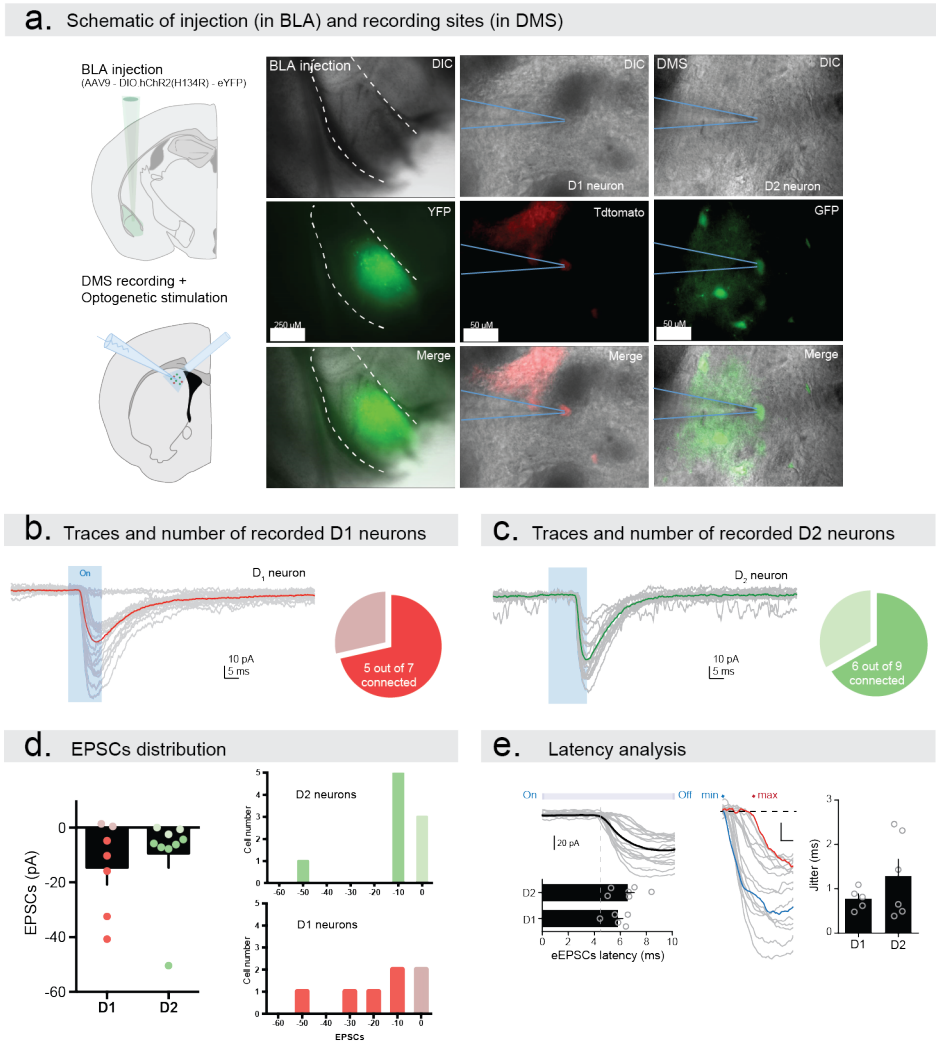

Figure 3.12 | Summary results of D1 and D2 neuronal responses to BLA stimulation in slice. (a) Scheme and images of the injection and recording sites. **(b)** Trace of the neuron with the biggest response amplitude for D1 neurons, and relative population of responsive neurons. **(c)** Trace of the neuron with the biggest response amplitude for D2 neurons, and relative population of responsive neurons. **(d)** EPSC to the first pulse of 20 trains of 14 Hz stimulation for both populations of neurons. **(e)** Latencies recorded for D1 and D2 responsive neurons.

The average latency to respond after stimulation was below 10 ms (**Figure 3.12e**), which indicates a direct connection between BLA and both populations of DMS (unlikely multisynaptic, which would take longer). More analysis will need to be performed to explore the potential differences in these synapses, as well as different potentiation mechanisms that might play a role in the circuitry.

Taken together, these results show that stimulation of the BLA to DMS projecting neurons is reinforcing, and that the instrumental action being reinforced has the hallmarks of a goal-directed behaviour. Additionally, these BLA excitatory projections impinge in both both D1 and D2 neurons in DMS.

DISCUSSION

In this study we analysed the sufficiency of BLA neurons projecting to DMS for the reinforcement of a new action. BLA has previously been implicated in tracking the value of specific stimuli associated with learning. Lesions in this area, as well as disconnections between BLA and DMS, lead to animals that are insensitive to changes in the value of an outcome associated with a learned action. To further study the role of the BLA neurons that project to DMS (BLA-DMS neurons), we trained animals in a self-paced, self-stimulation paradigm, where an action (lever pressing) leads to activations of these neurons. In WT the viral infection spread across different amygdala regions. We observed that animals receiving stimulations of amygdala-DMS neurons rapidly acquired high rates of lever pressing for self-stimulation. The learned behaviour was dependent on the contingency between the action and the outcome (as assessed by a session

of contingency degradation). Furthermore, stimulated animals could dynamically follow the stimulation, switching promptly between levers when both contingencies were reversed, and also when a block-training schedule was introduced. Therefore, the behaviour of the animals has the hallmarks of a goal-directed action.

Various works in diverse species has confirmed the role of DMS in goal-directed actions. The self-stimulation results confirm the role of amygdala-DMS projections in learning of a new goal-directed task.

As mentioned previously, in WT animals, the injection site is spread in areas around BLA, which could lead to potentially confounding effects of other projections to DMS from areas around BLA. However, retrograde tracers injected in DMS show very sparse labelling of the areas around BLA, while labelling this structure (Hart et al., 2014). Nevertheless, we confirm the results in a restrict number of transgenic animals that expressed Cre recombinase rather selectively in the BLA region of the amygdala (compared to neighbouring structures). Stimulation of BLA-to-DMS neurons in these Cre line animals also led to reinforcement of a novel action that could dynamically follow stimulation, thus confirming the results observed with the WT animals.

Still, the confirmation that action reinforcement is a direct result of stimulating projections from BLA to DMS is not possible in this set of experiments. Because of potential effects of back-propagation of action potentials to amygdala resulting from the optogenetic stimulation in DMS, our stimulation of projecting amygdala neurons could be stimulating collaterals targeting other areas (e.g. NAc, OFC). To test more specifically the circuitry involved in this reinforcement, one possibility would be to inhibit amygdala cell bodies while stimulating terminals in DMS. This

option, however, would still not discard potential effects of activation of collaterals that branch after the axons leave BLA. Another alternative would be to inhibit DMS terminals with DREADDs or muscimol during optogenetic stimulation of projection from BLA to DMS. If the effect of stimulation observed in this study is due to the specific projection to DMS, local inactivation of this efferent area should reduce behaviour output. Conversely, if the effect of stimulation its due to back-propagation leading to stimulation of other brain regions, DMS inactivation should have no effect on the observed behaviour.

As mentioned before, both BLA and CeA appear to be part of the parallel loops responsible for goal-directed and habitual actions. It is still not clear, however, how these areas behave and interact during instrumental learning. Interestingly, Namburi et al., 2015, observed that inactivation of BLA projecting to CeA neurons not only led to a decrease in freezing, but also to an increase in conditioned reward seeking. Therefore, it is also possible that this systems is not only regulating/modulating behaviour through their down-stream targets, but also that direct intra-amygdalar projection play a role in this balance.

Moreover, to extend our understanding about the dynamics of this system, it is essential to measure their activity in vivo during instrumental learning.

To the best of our knowledge, this study shows for the first time that DMS projecting BLA neurons can support the learning of novel action-outcome contingenies, in a flexible and dynamic manner that mimics a goal-directed strategy. These findings have important implications in our understanding of value association in action learning.

MATERIALS AND METHODS

Animals

All procedures were reviewed and performed in accordance with the Champalimaud Centre for the Unknown Ethics Committee guidelines, and approved by the Portuguese Veterinary General Board (Direcção Geral de Veterinária, approval 0421/000/000/2014). For the non-Cre experiments, Black C57BL male mice between 2 and 5 months of age were used in this study. Experiments were performed on the light cycle. For the Cre line experiments and line characterization, Gensat NL189 male mice between 2 and 5 months of age were used in this study (http://www.gensat.org/ShowFounderLineImages.jsp?gensatFounderLine=NL189-CRE). For the slice physiology experiments, Drd1-tdTtomato and Drd2-eGFP animals were crosses with NL189 animals. Mice between 2 and 3 months of age were used (Ade et al., 2011; Gong et al., 2003; Nelson et al., 2012). After surgery mice were housed individually under a 12 hours light/dark cycle.

Surgery and Histology

Surgeries were performed under anesthesia using a mix of oxygen (1 – 1.5 l/min) and isoflurane (1 – 3 %). For the behaviour and slice physiology experiments, each animal was bilaterally injected with 0.5 µl of viral solution in basolateral amygdala (anterior-posterior: 1.5 mm from bregma, mediolateral: 3.0 mm from bregma; dorsoventral: 4.0 mm from the brain surface) (Paxinos, G and Franklin, 2001), using a glass pipette, by pressure (nanojet II from Drummond Scientific, with 4.6 nl pulses at a rate of 0.2 Hz). The viruses injected were AAV2/9.CamKIIa.hChR.eYFP (University of North Carolina, titer 1.27 x 10^{13}) for ChR animals, and

AAV2/5.CamKIIa.hChR.eYFP (University of North Carolina, titer 6 x 10^{12}) for control animals in the non Cre dependent experiments, and AAV2/9.EF1a.DIO.hChR.eYFP (University of North Carolina, titer 1.69 x 10^{13}) for ChR animals, and AAV2/1.EF1a.DIO.eYFP (University of North Carolina, titer 1.85 x 10^{12}) for the Cre experiments. For optical stimuli delivery, fiber optics (200 µl diameter, NA=0.22) were implanted in DMS (anterior-posterior: 0.0 mm from bregma, mediolateral: 1.7 mm from bregma; dorsoventral: 2.0 mm from the brain surface) (Paxinos, G and Franklin, 2001). Animals for lever pressing behaviour were sacrificed after completion of the behaviour. Animals for slice recording were sacrificed 4 weeks after injection. Following anaesthesia, both control and ChR groups were perfused with saline and paraformaldehyde (4%). Their brains were removed for histological analysis and sectioned in 50 µm coronal slices (Leica vibratome). Both placement of fibers and spread of injection were investigated using a Zeiss AxioImager.M2 widefield fluorescence scanning microscope.

Behavioral procedures

2 weeks after surgery, the behaviour of the animals was tested in an instrumental task. Training took place in behavioural chambers (MED-PC, dimensions 23 cm x 20 cm x 19.5 cm – W x D x H) placed in sound attenuating boxes. Each chamber was equipped with a food magazine, a house light place on the wall on the left of the magazine and two retractable levers, one on each side of the magazine. MED-PC IV software was used to control the equipment and record lever presses, head entries to the magazine and laser on-set. Master8 software was used to drive the laser pulses, and Labview was used to record the behaviour of the animals in video. Optical stimuli were delivered to both ChR and control animals

with implantables fibers connect to a rotatory joint (Doric Lenses) coupled to a 200 mW and 437 nm laser (Shanghai Dream Lasers Technology Co., Ltd). Each stimulus was presented in 10 ms pulses of 14 Hz during 2 seconds, driven by an acousto-optic modulator (AA Opto Electronic) receiving TTL pulses from a Master8 stimulator (A.M.P.I.). The power of the laser was adjusted ex-vivo to be 5-10 mW per hemisphere at the tip of a reference fiber.

Acquisition train. During training, a session started with the illumination of the house light and extension of both levers. One lever was the active lever (AL) and one was the inactive lever (IL). The levers used as AL and IL were counterbalanced within groups. Optical stimuli to the dorsomedial striatum (DMS) were delivered contingently upon pressing the AL. Animals were trained on one session a day, and each session lasted 60 minutes with no maximum number of reinforcers. The wild-type animals were first trained under a CRF schedule. In the session after each animal earned a cumulative number of reinforcers bigger than 400, they progress to one day of fixed ratio 2 (FR2), where every to presses led to the delivery of stimulation. If their pressing rate dropped on the FR2 day, the animal was kept on that schedule for another 3 days (one animals out of 6). If their pressing rate increased, they progressed to 3 days of fixed ratio 4 (FR4). NL animals were trained on 11 days (i.e., the duration acquisition) of CRF.

Contingency degradation. After acquisition mice received contingency degradation training. In each CD session (60 minutes long) laser onset happened at a random time schedule and non-contingent upon lever press, i.e. independent of the animals' behaviour. The number of laser stimulations was yoked independently for animals, based on their average presses on the last 3 days of training. Following CD animals received a session of reinstatement, equal to the CRF training sessions before CD.

Switch Train. After the reinstatement day, wild-type animals had 3 days of switch training, were the previously active lever was inactive (no effect upon pressing) and the previously inactive lever was now the active lever (pressing leading to stimulation). The switch training consisted of 1 day of CRF, 1 day of FR2 and 1 day of FR4 (or in the case of the animal whose training did not progress to FR4 during acquisition, 1 day of CRF and 2 days of FR2). NL189 animals had 3 days of CRF in the switch training.

Block train. Finally, after switch train animals had 3 sessions of block train. During block sessions, the active and the inactive lever switched within sessions, dependent on number of reinforcers earned by the animal. In a x-reinforcer block train, the active and the inactive lever changed after x reinforcers were earned. The number of reinforcers per block was 50 or 100, dependent on the lever-pressing rate of the animals before the start of the block training, in order to insure that they could perform at least 4 blocks per session. Animals were trained for a minimum of 3 days, and got an extra day if their percentage of corrected presses did not reach 70%.

Slice electrophysiology

Standard procedures were used to prepare 300 μM thick coronal slices from 8- to 12-week-old NL-189 mice crossed with D1-tdtomato$^{(+)}$ or D2-GFP$^{(+)}$. The brain was dissected in ice-cold artificial cerebrospinal fluid (ACSF), mounted on an agar block, and sliced with a vibrating-blade microtome (HM 650 V, Carl Zeiss, Jena, Germany) at 4°C. Slices were maintained for 45 min at 37°C in an interface chamber containing ACSF equilibrated with 95% O_2/5% CO_2 and containing the following (in mM): 124 NaCl, 2.7 KCl, 2 $CaCl_2$, 1.3 $MgCl_2$, 26 $NaHCO_3$, 0.4 NaH_2PO_4, 18 glucose, 4 ascorbate. Recordings were performed with ACSF in a recording chamber at a temperature of 35°C. Neurons were visually identified with

infrared video microscopy using an upright microscope equipped with a 40X objective (Olympus, Tokyo, Japan). Patch electrodes (3–5 MΩ) were pulled from borosilicate glass tubing. For voltage clamp experiments to record evoked excitatory post-synaptic currents (eEPSCs), patch electrodes were filled with a solution containing the following (in mM): 120 K-gluconate, 20 KCl, 10 HEPES, 10 phosphocreatine, 4 Mg-ATP, and 0.3 Na-GTP (pH adjusted to 7.25 with KOH, respectively, 295 mOsm). D1 or D2 neurons were visualized using a fluorencesce lamp (X-cite) and GFP and Td-tomato filter. Neurons were clamped at a voltage of -70 mV. Evoked EPSCs were elicited 20 times by 14 Hz blue-light stimulation of the basolateral-amygdala local axon terminals of define neuronal populations expressing channelrodopsin in dorso-medial striatum. The eEPSCs shown in the figures were evoked by the first stimulation pulse of the 14 Hz train. Successful connections were scored if the amplitude of eIPSCs was higher than 4 pA, with the latency within 10 ms. Whole cell patch-clamp recordings were excluded if the access resistance exceeded 13 MΩ and changed more than 20% during the recordings. Data were recorded with a MultiClamp 700B (Molecular Devices) amplifier digitised at 10 kHz. Data were acquired and analysed with Clampex 10.0 and Clampfit 10.0, respectively (Molecular Devices). All chemicals for the internal and external solutions were purchased from Fluka/Sigma.

Statistical analysis

Statistical analyses were performed using GraphPad Prism 6 (GraphPad Software Inc., CA, USA). Repeated measures ANOVA were used to evaluate acquisition of lever presses and contingency degradation, followed by post hoc analyses using the Dunnet's test and the Sidák

correction when appropriate. Statistical significance was set at alpha=0.05. Results were represented as mean ± SEM.

ACKNOWLEDGEMENTS

We thank A. Vaz for colony management. This research was supported by the INDP Graduate Programme, a FCT fellowship to A.M.V., a Marie Curie to P.B. and an E.R.C. starting grant to R.M.C.

AUTHOR CONTRIBUTIONS

A.M.V. performed the optogenetical behavioural experiments and analyses. P.B. performed the slice physiology experiments and analyses. A.M.V. and R.M.C designed the experiments.

REFERENCES

Ade, K.K., Wan, Y., Chen, M., Gloss, B., and Calakos, N. (2011). An Improved BAC Transgenic Fluorescent Reporter Line for Sensitive and Specific Identification of Striatonigral Medium Spiny Neurons. Front. Syst. Neurosci. *5*, 32.

Balleine, B.W., Killcross, a S., and Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. J. Neurosci. *23*, 666–675.

Corbit, L.H., Leung, B.K., and Balleine, B.W. (2013). The role of the amygdala-striatal pathway in the acquisition and performance of goal-directed instrumental actions. J Neurosci *33*, 17682–17690.

Dickinson, A. (1985). Actions and Habits: The Development of Behavioural Autonomy. Philos. Trans. R. Soc. B Biol. Sci. *308*, 67–78.

Everitt, B.J., Cardinal, R.N., Parkinson, J.A., and Robbins, T.W. (2003). Appetitive behavior: impact of amygdala-dependent mechanisms of emotional learning. Ann. N. Y. Acad. Sci. *985*, 233–250.

Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E., et al. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature *425*, 917–925.

Hart, G., Leung, B.K., and Balleine, B.W. (2014). Dorsal and ventral streams: The distinct role of striatal subregions in the acquisition and performance of goal-directed actions. Neurobiol. Learn. Mem. *108*, 104–118.

Hatfield, T., Han, J.S., Conley, M., and Gallagher, M. (1996). Neurotoxic Lesions of Basolateral, But Not Central, Amygdala Interfere with Pavlovian Second-Order Conditioning and Reinforcer Devaluation Effects. J. Neurosci. *16*, 5256–5265.

Janak, P.H., and Tye, K.M. (2015). From circuits to behaviour in the amygdala.

Johansen, J.P., Cain, C.K., Ostroff, L.E., and Ledoux, J.E. (2011). Molecular mechanisms of fear learning and memory. Cell *147*, 509–524.

Kelley, A.E., Domesick, V.B., and Nauta, W.J.H. (1982). The amygdalostriatal projection in the rat-an anatomical study by anterograde and retrograde tracing methods. Neuroscience *7*.

LeDoux, J. (2007). The amygdala. Curr. Biol. *17*, 868–874.

Lingawi, N.W., and Balleine, B.W. (2012). Amygdala Central Nucleus Interacts with Dorsolateral Striatum to Regulate the Acquisition of Habits. J. Neurosci. *32*, 1073–1081.

Namburi, P., Beyeler, A., Yorozu, S., Calhoon, G.G., Halbert, S.A., Wichmann, R., Holden, S.S., Mertens, K.L., Anahtar, M., Felix-Ortiz, A.C., et al. (2015). A circuit mechanism for differentiating positive and negative associations. Nature *520*, 675–678.

Nelson, a. B., Hang, G.B., Grueter, B. a., Pascoli, V., Luscher, C., Malenka, R.C., Kreitzer, a. C., Lüscher, C., Malenka, R.C., and Kreitzer, a. C. (2012). A Comparison of Striatal-Dependent Behaviors in Wild-Type and Hemizygous Drd1a and Drd2 BAC Transgenic Mice. J. Neurosci. *32*, 9119–9123.

Ostlund, S.B., and Balleine, B.W. (2008). Differential involvement of the basolateral amygdala and mediodorsal thalamus in instrumental action selection. *28*, 4398–4405.

Paton, J.J., Belova, M.A., Morrison, S.E., and Salzman, C.D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature *439*, 865–870.

Paxinos, G and Franklin, K.B.J. (2001). Mouse Brain in Stereotaxic Coordinates.

Popescu, A.T., Popa, D., and Paré, D. (2009). Coherent gamma oscillations couple the amygdala and striatum during learning. Nat. Neurosci. *12*, 801–807.

Sah, P., Faber, E.S.L., Lopez De Armentia, M., and Power, J. (2003). The amygdaloid complex: anatomy and physiology. Physiol. Rev. *83*, 803–834.

Schoenbaum, G., Chiba, a a, and Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nat. Neurosci. *1*, 155–159.

Wall, N.R., De La Parra, M., Callaway, E.M., and Kreitzer, A.C. (2013). Differential innervation of direct- and indirect-pathway striatal projection neurons. Neuron *79*, 347–360.

4

DIRECT AND INDIRECT DORSOLATERAL STRIATUM PATHWAYS REINFORCE DIFFERENT ACTION STRATEGIES

‘The world unwraps itself to you, again and again,
as soon and you are ready to see it anew.”

Gregory Maguire

SUMMARY

It has been proposed that striatal output pathways have opposing roles in action reinforcement, with direct striatonigral neurons supporting positive reinforcement, and indirect striatopallidal neurons action avoidance. We uncovered that self-stimulation of either pathway in dorsolateral striatum (DLS) leads to positive reinforcement, but supports different action strategies. Activation of striatonigral neurons produced rapid action-specific reinforcement, while striatopallidal neuron self-stimulation resulted in generalization to similar actions, and less sensitivity to action-stimulation contingency.

INTRODUCTION

The basal ganglia, and the dorsal striatum in particular, are critical for action reinforcement (Albin et al., 1989; DeLong, 1990; Mink, 1996; Yin and Knowlton, 2006). The dorsal striatum, which can be further subdivided into dorsomedial (DMS) and dorsolateral (DLS) striatum, is mainly composed of two subpopulations of striatal medium spiny projection neurons (MSNs): dopamine D_1 receptor-expressing MSNs that reach directly the basal ganglia output nuclei and constitute the striatonigral or direct pathway (dMSNs); and dopamine D_2 receptors-expressing MSNs that constitute the striatopallidal or indirect pathway (iMSNs) (Gerfen et al., 1990). It has been suggested that striatonigral and striatopallidal neurons have opposing roles in reinforcement, with striatonigral neurons being important to learn positive reinforcement and indirect pathway neurons to learn to avoid undesired actions (Go/No-Go)

(Frank et al., 2004). Consistently, it has been shown that optogenetic self-stimulation of striatonigral neurons in DMS leads to reinforcement of actions that lead to stimulation, while self-stimulation of striatopallidal neurons leads to avoidance of actions that lead to stimulation (Kravitz et al., 2012). However, in DLS, which has been implicated in the consolidation of well-trained actions and in habit formation (Yin and Knowlton, 2006; Yin et al., 2009), both projection pathways are active during lever pressing for reward (Cui et al., 2013; Jin et al., 2014). Furthermore, extensive skill training leads to long-lasting potentiation of glutamatergic inputs into both d- and iMSNs (Yin et al., 2009). It has also been shown that striatal-specific deletion of $A_{2A}R$, which abolishes long-term potentiation onto iMSNs, impairs habit formation (Yu et al., 2009). These data suggest that in DLS both d- and iMSNs are involved in action learning and positive reinforcement, but support different action strategies.

RESULTS

Stimulation of both striatonigral and striatopallidal DLS neurons is reinforcing.

To investigate the role of DLS striatonigral and striatopallidal neurons in action reinforcement, we used a self-stimulation paradigm where we activated specifically each pathway upon lever pressing. We injected a virus expressing Channelrhodopsin-2 (ChR, AAV2/1) in a Cre-dependent manner into DLS of mice expressing Cre recombinase in either striatonigral (D_1-Cre, line EY217) (Gong et al., 2007) or striatopallidal neurons (D_2-Cre, line ER43) (Gong et al., 2007) (**Figure 4.1a**). Two weeks after infection, animals were trained in an operant box with two levers (**Figure 4.1b**): an

active lever where pressing resulted in the delivery of blue light (473 nm) into DLS, and an inactive lever (no light delivered). Self-paced reinforced lever presses resulted in the delivery of 10 pulses of light (for 2 sec, at 5Hz, 10ms wide pulses).

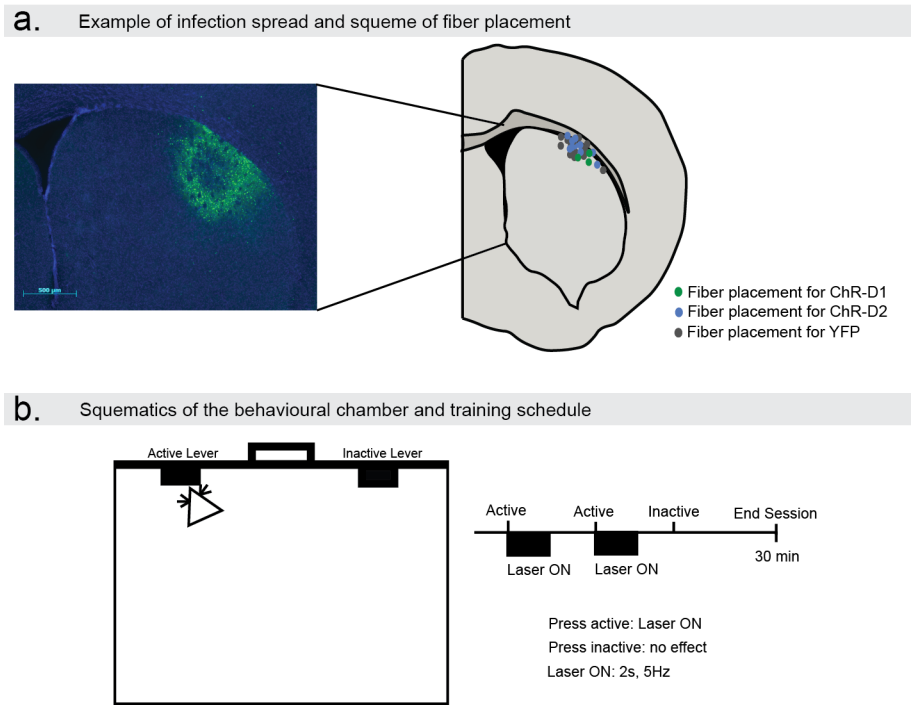

Figure 4.1 | Optogenetic self-stimulation of striatonigral and striatopallidal DLS neurons task. (a) Schematics and representative histology slice of injection and fiber placement sites in DLS. **(b)** Schematics of the operant box and the behavioural paradigm.

This stimulation frequency was chosen because it is similar to the endogenous activity of MSNs (Jin et al., 2014; Tecuapetla et al., 2014) and because continuous stimulation leads to a inhibition (or less excitation) during stimulation, following an initial peak of activity (**Figure 4.2a,b**).

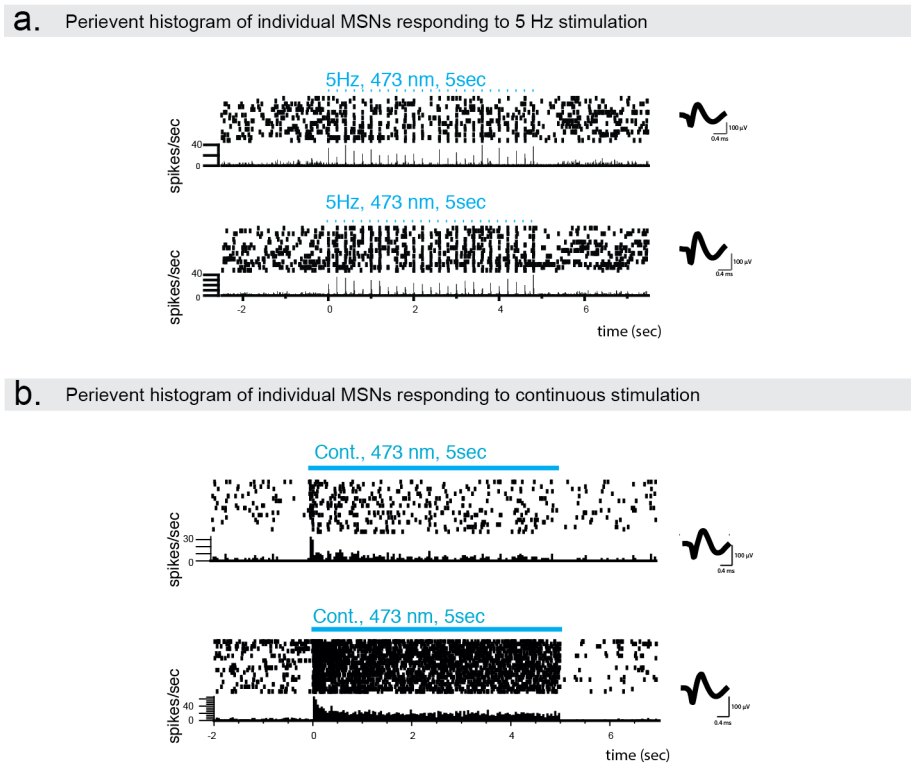

Figure 4.2 | Examples of peri-event histograms from individual MSN neurons expressing channelrodhopsin recorded in vivo and aligned to the on-set of a 473 nm blue laser stimulation. (a) stimulation with pulses of 5Hz, during 5 seconds, results in further biasing of ongoing activity (b) continuous light stimulation during 5 seconds leads to an initial peak of activation followed by less activation. We therefore chose to use 5Hz stimulation.

Each session lasted 30 minutes with no maximum number of reinforcers. Both groups of ChR-expressing mice increased the number of presses with training, and pressed significantly more than YFP controls (**Figure 4.3,** main effect of D_1 training $F_{14,140}=4.987$, $P<0.0001$; ChR effect $F_{1,10}=20.67$, $P=0.0011$; Interaction $F_{14,140}=4.883$, $P<0.0001$ (left panel); main effect of D_2 training $F_{31,527}=1.120$, $P=0.3026$; ChR effect $F_{1,17}=5.845$, $P=0.0271$; Interaction $F_{31,527}=1.505$, $P=0.0411$ (right panel)).

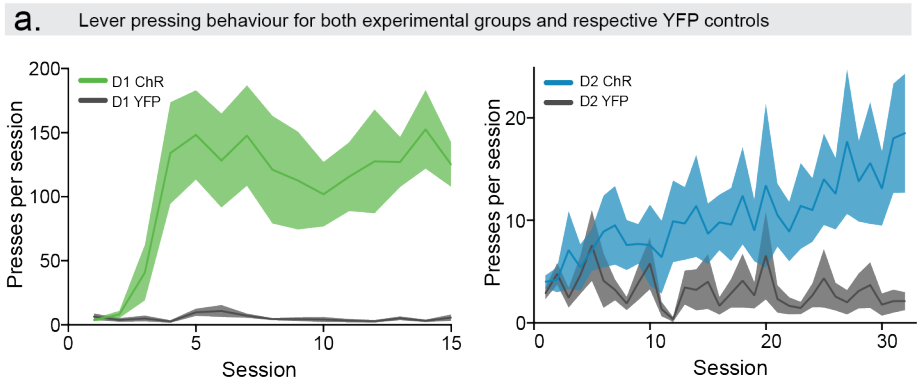

Figure 4.3 | Optogenetic self-stimulation of striatonigral and striatopallidal DLS neurons supports the reinforcement of actions. (a) Acquisition of lever pressing for ChR D_1-Cre animals infected animals (green line, n=6) and YFP controls (grey line, n=6) (left panel) and for ChR D_2-Cre animals infected animals (blue line, n=10) and YFP controls (grey line, n=9). Mean ± s.e.m plotted; * denotes p<0.05.

Distribution of lever pressing is different in D1 and D2 animals. Consistent with previous studies, D_1-Cre animals acquired lever pressing rather rapidly, and pressed the active lever significantly more than the inactive lever (**Figure 4.4a**, main effect of training $F_{14,280}$=5.143, P<0.0001; Lever and ChR effect $F_{3,20}$=21.21, P<0.0001; Interaction $F_{42,280}$=4.760, P<0.0001 (left panel); Main effect of training $F_{1,20}$=53.18, P<0.0001; lever and ChR effect $F_{3,20}$=45.38, P<0.0001; Interaction $F_{3,20}$=50.14, P<0.0001. Post hoc ChR active first day versus ChR active last day: P<0.0001 (right panel)). On the other hand, D_2-Cre animals expressing ChR were slower in acquisition, and showed a significant increase in lever pressing for both the active and the inactive levers (**Figure 4.4b**, main effect of D_2 training $F_{31,1054}$=1.516, P=0.0355; Lever and ChR effect $F_{3,34}$=3.111, P=0.0390; Interaction $F_{93,1054}$=1.093, P=0.2643 (left panel); Main effect of D_2 training $F_{1,34}$=8.282, P=0.0069; Lever and ChR effect $F_{3,34}$=3.858, P=0.0177; Interaction $F_{3,34}$=3.442, P=0.0274. Post hoc ChR active first day versus ChR

active last day: P<0.05; ChR inactive first day versus ChR inactive last day: P<0.05 (right panel)).

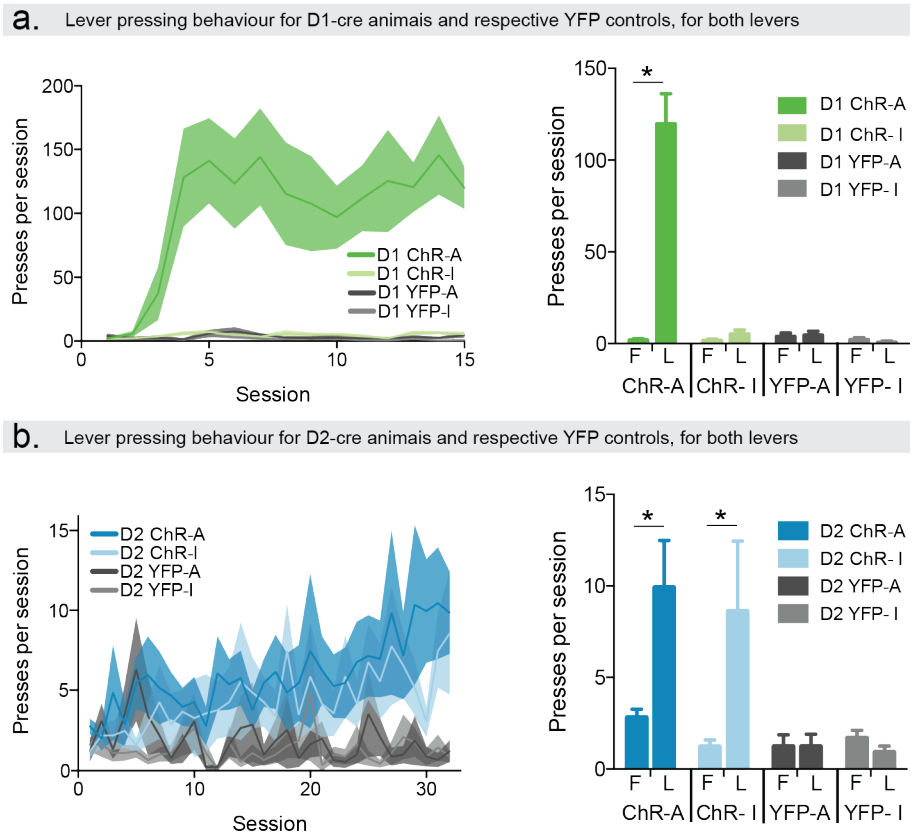

Figure 4.4 | Optogenetic self-stimulation of striatonigral and striatopallidal DLS neurons supports the reinforcement of different action strategies. **(a)** Acquisition of lever pressing in the active versus the inactive lever for ChR and YFP D1-Cre animals, throughout training (left panel) and in the first and last day of training (right panel). **(b)** Acquisition of lever pressing in the active versus the inactive lever for ChR and YFP D2-Cre animals, throughout training (left panel) and in the first and last day of training (right panel). Mean ± s.e.m plotted; * denotes p<0.05.

These data suggest that stimulation of both d- and iMSNs in DLS is reinforcing and not aversive, but leads to the development of different action strategies.

This does not stem from different number of pairings between action and reinforcer in D1- and D2-cre animals, because the same result was observed when matching the number of reinforcers between groups (**Figure 4.5a,b**).

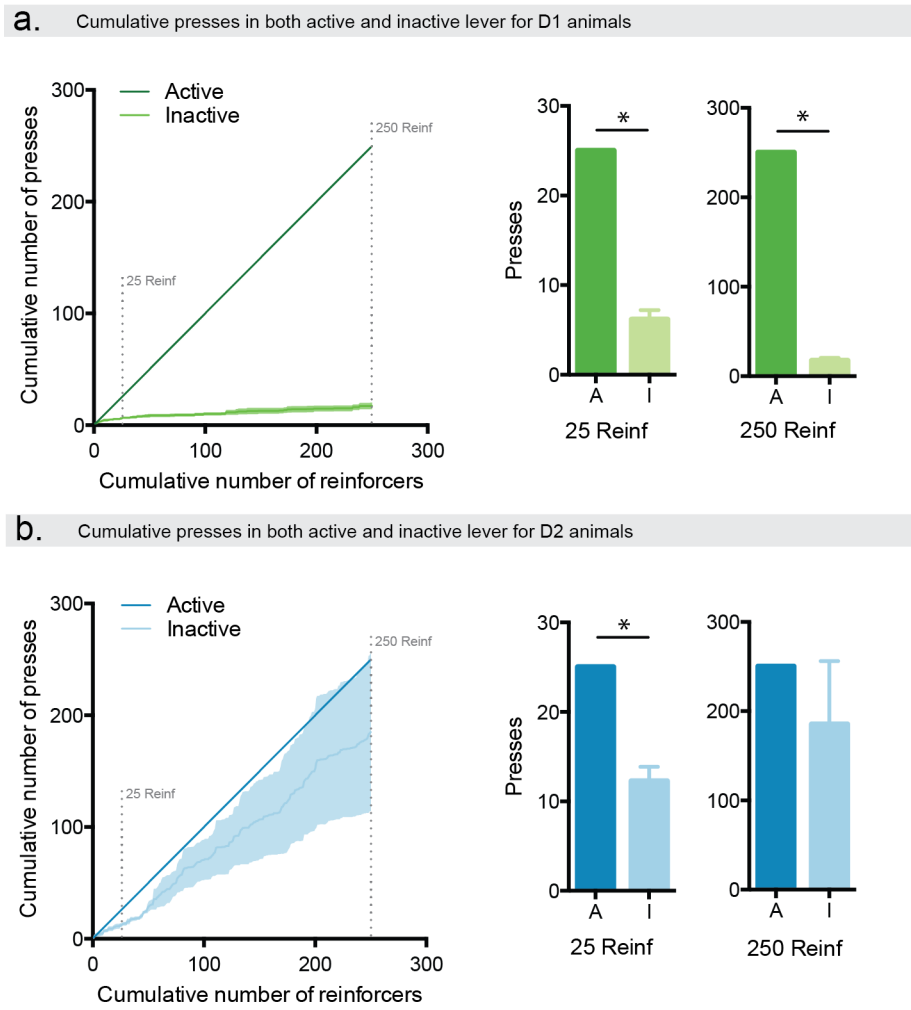

Figure 4.5 | Matching of both groups for the number of reinforcers highlights the different pressing distributions between active and inactive levers. **(a)** Cumulative active and inactive presses for D1-cre **(a)** and D2-cre **(b)** animals for the first 250 reinforcers earned (left panel). Comparison between the cumulative number of presses in the active and the inactive levers, for 25 and 250 reinforcers, for D1-cre **(a)** and D2-cre **(b)** animals (right panel). Mean ± s.e.m plotted; * denotes $p<0.05$.

To better characterize this dichotomy, we calculated the probability of pressing the active versus the inactive lever in every session. D_1-Cre animals expressing ChR showed a steady increase in the probability of pressing the active lever with training versus the probability of pressing the inactive lever (Main effect of training $F_{14,140}=3.447x10^{-14}$, $P>0.9999$; lever effect $F_{1,10}=688.3$, $P<0.0001$; Interaction $F_{14,140}=7.367$, $P<0.0001$. Post hoc p(active) versus p(inactive): $P<0.0001$ sessions 3-15, **Figure 4.6a**). On the other hand, D_2-Cre animals also had a higher probability of pressing the active than the inactive lever (Main effect of training $F_{31,558}=5.904x10^{-15}$, $P>0.9999$; Lever effect $F_{1,18}=6.961$, $P=0.0167$, **Figure 4.6c**) but this was mainly due to differences early in training (interaction $F_{\text{LeverxTraining time}}$ $F_{31,558}=1.903$), and eventually converged to a similar probability of pressing either lever (Posthocs not different for last days). To further investigate if this equal pressing of both levers resulted from generalization of lever pressing, or from avoidance of the active lever and shifting to the inactive lever after an active press, we calculated the probability of transition from an active stimulated lever press to a subsequent active press (or conversely, to an inactive press, **Figure 4.6b,d**).

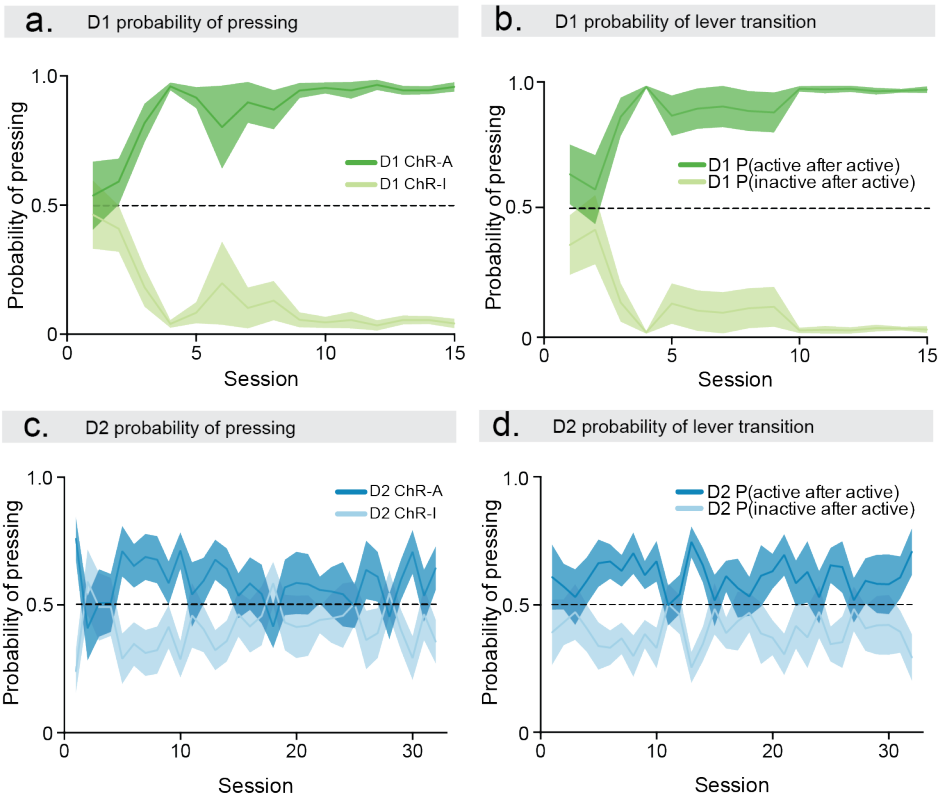

Figure 4.6 | D2-cre animals are not avoiding the active lever nor switch to an inactive press after an active one. (a) Probability of pressing the active versus the inactive lever for D1-Cre animals. **(b)** Probability of pressing the active versus the inactive lever for D2-Cre animals. **(c)** Probability of transition from an active lever press to a subsequent active lever press (versus an inactive press) for ChR D1-Cre. (d) Probability of transition from an active lever press to a subsequent active lever press (versus an inactive press) for ChR D2-Cre animals. Mean ± s.e.m plotted; * denotes $p<0.05$.

By the end of training, D_1-Cre animals reached a very high probability of pressing the active lever again after a previous reinforced active press (Main effect of training $F_{14,140}=1.752x10^{-14}$, $P>0.9999$; lever effect $F_{1,10}=310.9$, $P<0.0001$; Interaction $F_{14,140}=7.485$, $P<0.0001$. Post hoc p(active after active) versus p(inactive after active): $P<0.0001$ sessions 3-15, **Figure 4.6b**). D_2-Cre animals presented a slight but significantly higher probability of pressing the active lever after a reinforced active press

throughout training (main effect of training $F_{31,558}=5.696x10^{-15}$, P>0.9999; Lever effect $F_{1,18}=13.38$, P=0.0018; Interaction $F_{31,558}=1.176$, P=0.2362, although close to chance, **Figure 4.6d**), indicating that D_2-Cre mice were not just shifting to the inactive lever after an active lever press and then shifting back.

Sensitivity to contingency degradation is different in D1 and D2 experimental groups.

The data above suggest that self-stimulation of iMSNs leads to generalization between both levers, which is consistent with a role of these neurons in habit formation rather than goal-directed actions (Hilario et al., 2007, 2012). To evaluate if the actions of both groups were goal-directed and therefore sensitive to changes in the contingency between action and outcome, or habitual and therefore less sensitive to changes in contingency, we performed a contingency degradation (CD) experiment. During contingency degradation, the light stimulation was delivered non-contingently upon lever pressing, with the same probability of reinforcement per unit of time as during training. Following CD, animals underwent contingency reinstatement, where pressing the active lever would again lead to the delivery of stimulation. D_1-Cre animals decreased the number of presses during the CD session (**Figure 4.7a**, Main effect of contingency degradation $F_{2,20}=6.410$, P=0.0071; lever effect $F_{1,10}=45.68$, P=0<0001; Interaction $F_{2,20}=5.687$, P=0.0111. Post hoc ChR active Last day versus ChR active CD: P<0.001), and resumed their lever pressing behaviour during reinstatement (ChR active CD versus ChR active R: P<0.01). D_2-Cre animals, on the other hand, presented no changes in the number of presses during CD (**Figure 4.7a,** main effect of contingency degradation $F_{2,36}=0.09552$, P=0.9091; lever effect $F_{1,18}=3.295$, P=0.0862;

Interaction $F_{2,36}$=1.331, P=0.2769), suggesting that the lever pressing in these animals was habitual.

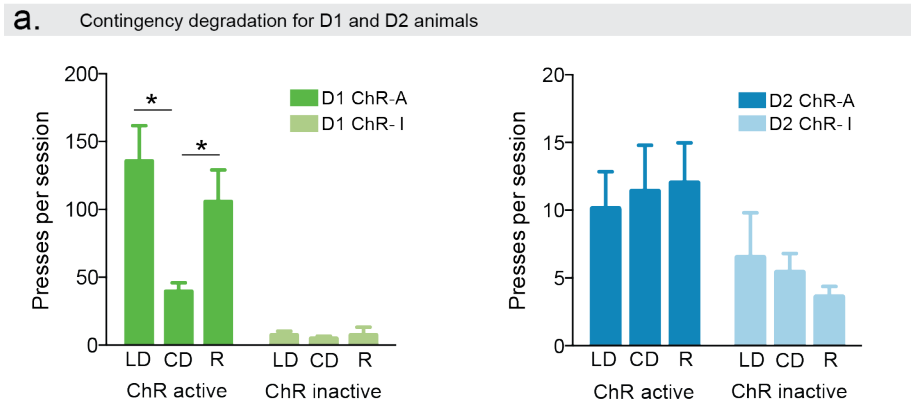

Figure 4.7 | Contingency degradation and reinstatement for D1-Cre (left) and D2-Cre (right) animals. Mean ± s.e.m plotted; A: active lever; I: inactive lever; LD: last day of training; CD: contingency degradation day; R: reinstatement day. * denotes p<0.05

The insensitivity to contingency degradation in iMSN-stimulated animals is unlikely to be due to a floor effect, since it has been previously shown that animals that press less tend to be more sensitive to contingency manipulation (Hilario et al., 2007).

DISCUSSION

In this study, we show that self-stimulation of both striatonigral and striatopallidal DLS neurons is sufficient to positively reinforce actions, but that stimulation of each pathway supports the learning of different action strategies. While animals receiving stimulation in striatonigral neurons acquired the task rapidly, pressed almost exclusively the active lever and were sensitive to changes in contingency, mice self-stimulating

striatopallidal neurons acquired lever pressing more slowly (and never pressed as much), generalized pressing from the active to the inactive lever, and were insensitive to contingency degradation.

These results suggest that pairing activation of striatonigral neurons in DLS with an action leads to the establishment of a goal-directed relation between that action and the outcome, while pairing activation of striatopallidal neurons in DLS with an action supports the formation of a stimulus-response habit that generalizes to similar manipulanda (Hilario et al., 2012) and is insensitive to changes in contingency (Yin and Knowlton, 2006).

These conclusions are consistent with the role of long-lasting plasticity of glutamatergic inputs into DLS striatopallidal neurons in both habit formation and skill consolidation (Yin et al., 2009; Yu et al., 2009). These data also raise the interesting possibility that DLS might not be homogenously involved in habit formation; direct and the indirect pathways in DLS could support different action strategies and compete for action control. This role may be different in DMS, where striatonigral and striatopallidal seem to support opposite roles in reinforcement (Kravitz et al., 2012) (**Figure 4.8a**). These results could also be consistent with a view in which both striatal projection pathways are involved in action selection, with striatonigral neurons supporting the execution of the desired action pattern, and striatopallidal neurons avoiding the execution of competing action patterns (Cui et al., 2013; Mink, 1996); in this view self-stimulation of striatopallidal neurons could mainly support the avoidance of actions other than lever pressing in that particular context (**Figure 4.8b**).

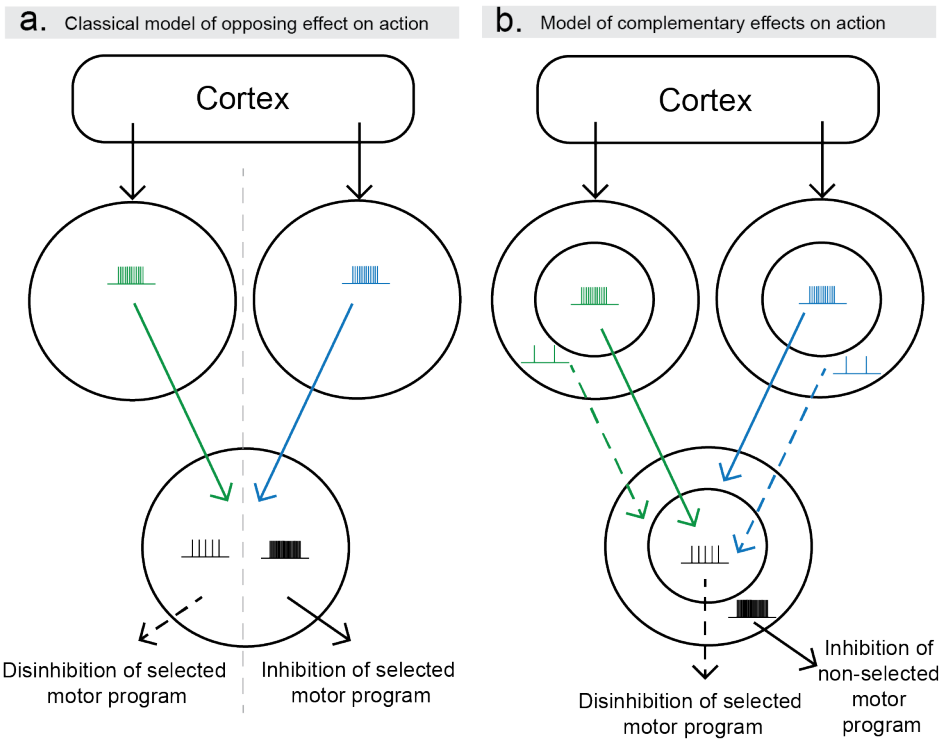

Figure 4.8 | Schematics of two different models of striatal output control of actions. (a) Classical model of the dichotomous effect of each pathway on action, where striatonigral pathway neuronal activation leads to the execution of the appropriate action, while activation of the striatopallidal pathway leads to inhibition of actions (Albin et al., 1989). **(b)** Schematics of a model of action output where both pathways are involved in action selection, with striatonigral activation leading to the execution of the desired action, and striatopallidal involved in inhibiting competing actions (Mink, 1996).

Interestingly, a recent paper further confirms this double involvement of both DLS pathways, and could help shape a possible alternative to the their role in DLS dynamics (O'Hare et al., 2016): ex-vivo DLS measures of trained animals showed a positive correlation between the amplitude of cortical evoked calcium transients (that relay activity information) in both dMSNs and iMSNs and degree of habitual control of the action. Additionally, the relative firing time of the two population also correlated with habitual control; dMSNs tended to fire before iMSNs in more habitual animals, but after in more goal-directed ones. This raises the intriguing but

estremely parsimonious hypothesis that the transition to a habit in DLS depends on a shift in timing between the two populations.

Still, it is clear from these results that in DLS, self-stimulation of striatopallidal neurons is not aversive. In this context, it is interesting to note that optogenetic stimulation of iMSNs leads to the activation of a subset of cortical M1 neurons (Oldenburg and Sabatini, 2015), and that inactivation of iMSNs does not necessarily increase basal ganglia output activity (Tecuapetla et al., 2014), further underscoring that the functional organization of basal ganglia output is more complex than classically proposed. Lever-pressing activity in both pathways precedes action initiation (Cui et al., 2013). Therefore, plasticity associated with instrumental learning could be occurring at recently active corticostriatal synapses (and could be different for d- and i-MSN synapses). Alternatively, stimulation of MSNs could specifically select inputs onto cortical neurons that were previously active through the cortico-basal ganglia-thalamocortical loop.

Taken together, these results show that in DLS both striatonigral and striatopallidal activation can support positive reinforcement of actions paired with that activation, but that the action strategies learned are different. These findings may have implications for understanding the basal ganglia circuitry underlying compulsive actions and persistent habits.

MATERIALS AND METHODS

Animals

All procedures were reviewed and performed in accordance with the Champalimaud Centre for the Unknown Ethics Committee guidelines, and

approved by the Portuguese Veterinary General Board (Direcção Geral de Veterinária, approval 0421/000/000/2014). Male mice between 2 and 5 months of age, resulting from the backcrossing of BAC transgenic mice into Black C57BL for at least 8 generations (which express the Cre recombinase under the control of the dopamine D1a (EY217 line) or D2 (ER43 line) receptor promoters) were used in this study. These lines were chosen because their expressions is more restricted to striatum, to avoid possible contaminations from any cortical stimulations. After surgery mice were housed individually under a 12 hours light/dark cycle. Experiments were performed on the light cycle.

Surgery and Histology

Surgeries were performed under anesthesia using a mix of oxygen (1 – 1.5 l/min) and isoflurane (1 – 3 %). Each animal was bilaterally injected with 1.5 µl of viral solution in dorsolateral striatum (DLS – anterior-posterior: 0.38 mm from bregma, mediolateral: 2.5 mm from bregma; dorsoventral: 2.2 mm from the brain surface)[16], using a glass pipette, by pressure (nanojet II from Drummond Scientific, with 4.6 nl pulses at a rate of 0.4 Hz). The viruses injected were AAV2/1.EF1a.DIO.hChR.eYFP (University of North Carolina, titer 5.58 x 10^{12}) for ChR animals, and AAV2/1.EF1a.DIO.eYFP (University of North Carolina, titer 1.85 x 10^{12}) for control animals. For optical stimuli delivery, fiber optics (200 µl diameter, NA=0.22)[17] were implanted at the site of injection, 2.0 mm from the brain surface. Animals were sacrificed after completion of the behaviour. Following anaesthesia, both control and ChR groups were perfused with saline and paraformaldehyde (4%). Their brains were removed for histological analysis and sectioned in 50 µm coronal slices (Leica vibratome). Both placement of fibers and spread of injection were

investigated using a Zeiss AxioImager.M2 wide field fluorescence-scanning microscope.

Behavioural procedures

2 weeks after surgery, the behaviour of the animals was tested in an instrumental task. Training took place in behavioural chambers (MED-PC, dimensions 23 cm x 20 cm x 19.5 cm – W x D x H) placed in sound attenuating boxes. Each chamber was equipped with a food magazine, a house light place on the wall on the left of the magazine and two retractable levers, one on each side of the magazine. MED-PC IV software was used to control the equipment and record lever presses, head entries to the magazine and laser on-set. Master8 software was used to drive the laser pulses, and Labview was used to record the behaviour of the animals in video. Optical stimuli were delivered to both ChR and control animals with implantables fibers[17] connect to a rotatory joint (Doric Lenses) coupled to a 200 mW and 437 nm laser (Shanghai Dream Lasers Technology Co., Ltd). Each stimulus was presented in 10 ms pulses of 5 Hz[18] during 2 seconds, driven by an acousto-optic modulator (AA Opto Electronic) receiving TTL pulses from a Master8 stimulator (A.M.P.I.). The power of the laser was adjusted ex-vivo to be 5-10 mW per hemisphere at the tip of a reference fiber. During training, a session started with the illumination of the house light and extension of both levers. One lever was the active lever (AL) and one was the inactive lever (IL). The levers used as AL and IL were counterbalanced within groups. Optical stimuli to the dorsolateral striatum (DLS) were delivered contingently upon pressing the AL. Animals were trained one session a day, during 30 minutes each and no limit on the number of stimuli, on a continuous reinforcement (CRF) schedule, where each press led to one stimulus. For D_2 animals, animals

were trained on CRF for 32 days. D_1 animals were trained for at least 15 days. After acquisition mice received contingency degradation training. In each CD session (30 minutes long) laser onset happened at a random time schedule and non-contingent upon lever press, i.e. independent of the animals' behaviour. The number of laser stimulations was yoked independently for each animal, based on their average presses on the last 3 days of training. The D_1 group had 1 session of CD, while the D_2 group had 2 session of CD (to guarantee that indeed they were not sensitive to CD). Following CD animals received a session of reinstatement, equal to the CRF training sessions before CD.

In Vivo Recordings

In vivo recordings were performed with movable bundle of 16 platinum-coated tungsten microwires with coupled guide cannulas to introduce a fibre optic 200–300 μm away from the tip of the electrodes (Innovative-Neurophysiology). Neural activity and light stimulation timestamps were recorded with a Cerebrus recording system (Blackrock Microsystems). After recording units were offline sorted (Offline Sorter, Plexon Inc.) and time stamps and waveforms were exported to MATLAB for further analyses.

Statistical analysis

Statistical analyses were performed using GraphPad Prism 6 (GraphPad Software Inc., CA, USA). Repeated measures ANOVA were used to evaluate acquisition of lever presses and contingency degradation, followed by post hoc analyses using the Dunnet's test and the Sidák correction when appropriate. Statistical significance was set at alpha=0.05. Results were represented as mean ± SEM.

ACKNOWLEDGMENTS

We would like to thank A. Vaz for mouse colony management, G. Fioreze for help in some experiments, and J. A. Silva for help in setting up the optogenetic experiments. This work was supported by fellowships to P.G.F, A.M.V and F.T. from Fundação para a Ciência e Tecnologia, and Grants from ERA-NET (F4T), European Research Council (COG 617142) and HHMI (IEC 55007415) to R.M.C.

AUTHOR CONTRIBUTION

A.M.V. and P.G.F. performed the optogenetical behavioural experiments and analyses. F.T. performed the electrophysiology experiments and analyses. A.M.V., P.G.F. and R.M.C designed the experiments.

REFERENCES

Albin, R.L., Young, A.B., and Penney, J.B. (1989). The functional anatomy of basal ganglia disorders. Trends Neurosci. *12*, 366–375.

Cui, G., Jun, S.B., Jin, X., Pham, M.D., Vogel, S.S., Lovinger, D.M., and Costa, R.M. (2013). Concurrent activation of striatal direct and indirect pathways during action initiation. Nature *494*, 238–242.

DeLong, M.R. (1990). Primate models of movement disorders of basal ganglia origin. Trends Neurosci. *13*, 281–285.

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

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*, 1429–1432.

Gong, S., Doughty, M., Harbaugh, C.R., Cummins, A., Hatten, M.E., Heintz, N., and Gerfen, C.R. (2007). Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. J. Neurosci. *27*, 9817–9823.

Hilario, M.R.F., Clouse, E., Yin, H.H., and Costa, R.M. (2007). Endocannabinoid signaling is critical for habit formation. Front. Integr. Neurosci. *1*, 1–12.

Hilario, M.R.F., Holloway, T., Jin, X., and Costa, R.M. (2012). Different dorsal striatum circuits mediate action discrimination and action generalization. Eur. J. Neurosci. *35*, 1105–1114.

Jin, X., Tecuapetla, F., and Costa, R.M. (2014). Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. Nat. Neurosci. *17*, 423–430.

Kravitz, A. V, Tye, L.D., and Kreitzer, A.C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat. Neurosci. *15*, 816–818.

Mink, J.W. (1996). the Basal Ganglia: Focused Selection and Inhibition of Competing Motor Programs. Prog. Neurobiol. *50*, 381–425.

O'Hare, J.K., Ade, K.K., Sukharnikova, T., Van Hooser, S.D., Palmeri, M.L., Yin, H.H., and Calakos, N. (2016). Pathway-Specific Striatal Substrates for Habitual Behavior. Neuron *89*, 472–479.

Oldenburg, I.A.A., and Sabatini, B.L.L. (2015). Antagonistic but Not Symmetric Regulation of Primary Motor Cortex by Basal Ganglia Direct and Indirect Pathways. Neuron *86*, 1174–1181.

Tecuapetla, F., Matias, S., Dugue, G.P., Mainen, Z.F., and Costa, R.M. (2014). Balanced activity in basal ganglia projection pathways is critical for contraversive movements. Nat. Commun. *5*, 4315.

Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. *7*, 464–476.

Yin, H.H., Mulcare, S.P., Hilario, M.R.F., Clouse, E., Holloway, T., Davis, M.I., Hansson, A.C., Lovinger, D.M., and Costa, R.M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nat. Neurosci. *12*, 333–341.

Yu, C., Gupta, J., Chen, J.-F., and Yin, H.H. (2009). Genetic deletion of A2A adenosine receptors in the striatum selectively impairs habit formation. J. Neurosci. *29*, 15100–15103.

5

DISCUSSION

"...And although my speculations greatly please myself, I believe that others have theirs, which perhaps please them still more."

René Descartes

In this thesis we focused on different circuits mediating the balance between goal-directed and habitual actions. Over the years a growing amount of research has unveiled different circuitry behind one or the other behavioural strategy (Corbit and Balleine, 2005; Coutureau and Killcross, 2003; Faure et al., 2005; Killcross and Coutureau, 2003; Yin et al., 2004, 2005). One of the most striking conclusions arising from work in these circuits is that lesions or inactivation do not lead to a complete loss of the action, but to a switch to the other behaviour strategy, whose circuits is still intact. Therefore, both seem to co-exist to some extent during learning and execution of the action. However, if one takes on the concept of plastic parallel pathways that compete to control behaviour, one has to wonder were that control and competition takes place. This question was one of the start points of this thesis.

In the first chapter of results (Chapter 2) we studied the role of the dopamine transporter (DAT) in the dynamic switch between goal-directed actions and habits, based on context dependent cues. We trained animals to press a lever in two separate contexts, under schedules that would either prone to goal-directed or to habitual behaviour (Gremel and Costa, 2013). We observed that animals with a heterozygous deletion in DAT were habitual in both schedules, while wild-type littermates were goal-directed in one and habitual in the other (as expected from previous results (Gremel and Costa, 2013; Hilario et al., 2007)). Further, we observed that the same mutation did not affect learning of a goal-directed association when animals were trained in one single context under a single schedules that prone to learning of an action-outcome associations.

In the second results' chapter (Chapter 3) we focused on the neurons projecting from the basolateral amygdala (BLA) to dorsomedial striatum (DMS). Animals expressing Channelrhodopsin in BLA were trained to press a lever to receive stimulation in DMS. We observed that animals learned to press to receive stimulation of the neurons that project from amygdala to DMS, and that self-stimulation is enough to induce a reinforcing effect. The lever pressing was sensitive to contingency degradation (one hallmark of goal-directed actions) and could very dynamically follow changes in contingency (seen by changing active and inactive levers, and by block training). In a restricted set of animals we confirmed that the amygdala cells responsible for this phenotype are likely projecting cells from the BLA. Finally, we also observed that, unlike previously reported (Wall et al., 2013), cell in BLA appear to project to both direct and indirect pathway neurons in DMS.

In the fourth chapter we studied the reinforcing role of both DLS D1 and D2 neurons in reinforcement. We trained mice to press a lever to receive stimulation of either pathway. We observed that although both stimulations led to action performance, the type of action reinforced appeared to be different. While D_1-cre animals rapidly increased pressing rate and were sensitive to contingency degradation, the D_2-cre animals had a reduced rate of pressing, similar to both the active (which delivered stimulation) and the inactive (no consequence of pressing) lever, and were insensitive to contingency degradation. While previous reports had shown iMSNs activation in DMS to promote aversive responses (Kravitz et al., 2012), at least in DLS we observed a different phenotype.

These systems are not, obviously, working in isolation from each other or from the environment. Although these projects were told in separate chapters, there are several overlapping features between these circuits. Further, in amygdala there is a pattern of parallel relationship between the control of goal-directed and habitual action that resembles corticostriato-thalamic loops observed in several experiments. While BLA appears to be important for goal-directed action control, in particular to the value update of the specific outcome associated to the action, CeA is necessary to learn stimulus-response actions (Corbit and Balleine, 2005). Since there are several areas involved in these loops, a pertinent question is how are they balanced each other to take control of the behaviour. One possibility is that there is a receiving downstream region that gates the control of the behaviour, sensing information from both circuits. Yet, there is the possibility that the system is directly competing, which means that somewhere along the loop (or even at several points of parallel processing) one structure is inhibiting the other, keeping it under control. Interestingly, and while not much is known about lateral connections between DMS and DLS, there are several studies that look at intra-amygdalar connections. Besides their inter-regions projects, there are several intra-amygdalar connections (Sah et al., 2003), with different central amygdala regions receiving projections from different areas of the BLA (for an in depth analysis, see the review by Sah, P *et al*, 2003). Furthermore, it is also relevant to note that the intercalate cell (IC) nucleus of amygdala, which are inhibitory, receive projections from the BLA and project to CeA (Marowsky et al., 2005). While medial ICs receive projections from BLA and project to CeA, lateral ICs project to projection cells in BLA. Aditionally, activation of BLA to central lateral (CeL) projections, which results in inhibition of the central medial amygdala (CeM), leads to an

anxiolytic effect (that is not seen with BLA activation) (Tye et al., 2011). BLA to CeM direct projection, on the other hand, are associated with fear conditioning (Namburi et al., 2015). This system provides not only what appear to be distinct circuits mediating appetitive and aversive learning, but also a pathway of control of action from a structure associated with goal-directed action (BLA) over a structure associated with habits (CeA).

There are also strong connections and relationships between the amygdala and the dopaminergic system. CeA projects to DA cells in the midbrain, but its projections are mainly to the lateral SNc, and minimal to VTA (Wallace et al., 1992; Zahm et al., 1999). This is the same midbrain area that is known to project to DLS. The majority of existing CeA to VTA projections do not target the DA neurons of this region (Wallace et al., 1992). Conversely, VTA and medial SNc project to both CeA and BLA. Lateral SNc projections, however, do not show any labelling in amygdala (Fallon and Moore, 1978; Swanson, 1982). Around 83% of the dendrites found to form synapses from DA neurons onto BLA were from putative projection neurons (Muller et al., 2009). While in both primate and rat BLA projection and inter neurons are depolarized and excited by D1 receptor activation, and DA can also increase the probability of glutamatergic release pre-synaptically (Muly et al., 2010; Kröner et al., 2005), the intercalated cells are inhibited by DA release sensed by D1 receptors (Marowsky et al., 2005). In BLA, D1 and D2 have been found to co-localize in the same neuron (Maltais et al., 2000). Both D1 and D2 antagonists blocked the increase in excitability in BLA neurons by DA. Furthermore, D1 agonist by itself increased excitability. This might mean that D1 increases excitability dependently on D2 (which can be tonic active even without further agonists) (Kröner et al., 2005). As in striatum,

activation of DA receptors can lead to a bidirectional plasticity system in BLA (Li and Rainnie, 2014).

There are complex and hard to interpret results that arise from these connections, especially since a plethora of instrumental task with different learning rules and parameters have been used to study them. Animals working for VTA stimulation have a dose dependent drop in pressing response upon muscimol infusion into BLA (Simmons et al., 2007). However, this task involves decreasing power of stimulation upon pressing, which can be reset by pressing another lever. The author also observed that animals showed a decrease in resetting the values to optimal. D1 antagonist infusion in either BLA or CeA impairs lever pressing learning (CeA more than BLA) (Andrzejewski et al., 2005). The same infusions have no effect in behaviour expression. D1 antagonist infused in caudal BLA during maintenance decreases cocaine-seeking behaviour, while agonist increased (Mashhoon et al., 2009).

Therefore, there is evidence that amygdala and the dopamine system have a tight cooperation in a variety of behaviours. Dopamine can modulate activity of BLA and CeA directly through its projections to those areas, but also indirectly through the IC. CeA, on the other hand, can also influence DA activity through its projections to SNc and VTA. There is, however, an extra possible stream of control. In the nucleus accumbens (NAc), electrical stimulation of BLA, but not CeA, results in an increase in DA signal (Floresco et al., 1998; Howland et al., 2002). This efflux of DA was independent from VTA inactivation, which indicates that projection from BLA to NAc can directly lead to an increase in DA release. Although it is not known if the same dynamics is present in DMS, the projections of BLA

and VTA to that area largely overlap (Kelley et al., 1982). Therefore, it is conceivable to speculate that this could be an extra possible layer of control over DA signal in DMS.

DA has had a lot as attention as an error prediction signal, due to its response to unpredictable US, that shifts with learning to a CS that predicts that stimulus (Schultz, 1998). However, DA neurons are known to respond to more than just CS predicting reward. In fact, DA neurons also respond to salient events, not just to unpredictable rewarded stimuli (reviewed in Bromberg-Martin et al., 2011; Horvitz, 2000). When glutamatergic inputs arrive to the striatum (from cortical, thalamic and amygdalar projections), DA is in a excellent position to selectively gate strong glutamatergic inputs to striatum (for a full description of how, check (Horvitz, 2002)), effectively selecting which "inputs" the striatum listens to. In accordance, lesions on DA terminals in DMS lead to animals that are insensitive to contingency degradation in a 2 action – 2 outcome RR schedule training (Lex and Hauber, 2010), while lesions on DA terminals in DLS lead to animals that are insensitive to outcome devaluation in a 2 action – 2 outcome variable interval training schedule (Faure et al., 2005).

One interesting observation from skill learning came from the observation of the effect of lesions in DLS on rotarod performance – although lesions in DMS affect performance in the early stages of skill learning, it has no effect on later stages. On the contrary, both early and late lesions in DLS affect skill performance. This could be related to our observation that in DLS, both D1 and D2 type neurons are involved in reinforcement. It is possible that from the beginning of instrumental learning, DLS is slowly contributing to learning, until the point where it becomes DMS

independent. In the 2 contexts – 2 schedules task, it was observed that there were neuronal populations in both DMS and DLS that modulated their response upon lever pressing to a specific context, or to both contexts (Gremel and Costa, 2013). While the relative percentage of neurons that respond to both contexts increases in DMS, it remains stable in DLS through training. In DMS, however, the modulation rate of DMS neurons that respond specifically to one context increases for the RR context, and decreases for the RI context. In DLS, on the other hand, although the relative percentage of neurons responding to both remains similar after 6 days of train, they form a binomial distribution, where neurons became more inhibited in the RR schedule and more activated in the RI schedule. Importantly, we do not have access to the type of neurons that were modulated. Therefore, it is not possible to establish the percentage of D1 and D2 neurons in that study (Gremel and Costa, 2013). Given the results of our study, it would be extremely interesting to follow the effects of learning in both neuronal populations.

It is also relevant to note that the circuit is a lot more complex than what is described on this thesis. As mentioned in Chapter 1, different cortical and thalamic regions innervate different dorsal striatal areas. Although these are outside the scoop of this study, their role in the balance between the two types of behaviour can't be overlooked

There are, obviously, several open questions. To begin with, the different reinforcing effects observed with stimulation of striatopallidal and striatonigral neurons in DLS and DMS. As discussed above, there are several possible explanation for why the stimulation of iMSNs appears to be appetitive in DLS, but aversive in DMS. The most obvious one being

that they are different areas. It is indeed possible that the role of this population is different in the two areas. However, it is also possible that the differences observed are due to differences in the protocol (of training and stimulation). Tying projects together, although BLA cells are thought to project mostly to D1 cells in DMS, we have shown evidences that they project to both populations. Therefore, it is conceivable that the reinforcing signal provided by BLA is reaching both neuronal populations. BLA could even be differentially lowering or increasing excitability of either cell type by its DA local release control.

To further advance these questions, a more specific look at the circuits needs to be performed. In particular, it would be of extreme interest to image BLA, dMSNs and iMSNs in DLS and DMS while animals are learning a lever-pressing task under a RR or a RI schedule. Measuring DA concentration in DLS and DMS during the performance of a similar task could also further advance the knowledge of this system. Finally, and to have a better gage on were the dynamic control over the systems is, it would be interesting to disconnect both DMS and DLS, and also BLA and CeA.

The work presented in this thesis contributes to further dissection of striatal sub-circuits, and of projections impinging on striatum, in gating learning/execution of goal-directed and habitual actions.

REFERENCES

Andrzejewski, M.E., Spencer, R.C., and Kelley, A.E. (2005). Instrumental learning, but not performance, requires dopamine D1-receptor activation in the amygdala. Neuroscience *135*, 335–345.

Bromberg-Martin, E.S., Matsumoto, M., and Hikosaka, O. (2011). Dopamine in motivational: rewarding, aversive,and alerting. Neuron *68*, 815–834.

Corbit, L.H., and Balleine, B.W. (2005). Double Dissociation of Basolateral and Central Amygdala Lesions on the General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer. J. Neurosci. *25*, 962–970.

Coutureau, E., and Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav. Brain Res. *146*, 167–174.

Fallon, J.H., and Moore, R.Y. (1978). Catecholamine innervation of the basal forebrain. IV. Topography of the dopamine projection to the basal forebrain and neostriatum. J. Comp. Neurol. *180*, 545–580.

Faure, A., Haberland, U., Conde, F., and El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. J. Neurosci. *25*, 2771–2780.

Floresco, S.B., Yang, C.R., Phillips, A.G., and Blaha, C.D. (1998). Basolateral amygdala stimulation evokes glutamate receptor-dependent dopamine efflux in the nucleus accumbens of the anaesthetized rat. Eur. J. Neurosci. *10*, 1241–1251.

Gremel, C.M., and Costa, R.M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nat. Commun. *4*, 2264.

Hilario, M.R.F., Clouse, E., Yin, H.H., and Costa, R.M. (2007). Endocannabinoid signaling is critical for habit formation. Front. Integr. Neurosci. *1*, 1–12.

Horvitz, J.C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non- reward events. Neuroscience *96*, 651–656.

Horvitz, J.C. (2002). Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum. Behav. Brain Res. *137*, 65–74.

Howland, J.G., Taepavarapruk, P., and Phillips, A.G. (2002). Glutamate receptor-dependent modulation of dopamine efflux in the nucleus accumbens by basolateral, but not central, nucleus of the amygdala in rats. J. Neurosci. *22*, 1137–1145.

Kelley, A.E., Domesick, V.B., and Nauta, W.J.H. (1982). The amygdalostriatal projection in the rat-an anatomical study by anterograde and retrograde tracing methods. Neuroscience *7*.

Killcross, S., and Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb. Cortex *13*, 400–408.

Kravitz, A. V, Tye, L.D., and Kreitzer, A.C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat. Neurosci. *15*, 816–818.

Kröner, S., Rosenkranz, J.A., Grace, A. a, and Barrionuevo, G. (2005). Dopamine modulates

excitability of basolateral amygdala neurons in vitro. J. Neurophysiol. *93*, 1598–1610.

Lex, B., and Hauber, W. (2010). The role of dopamine in the prelimbic cortex and the dorsomedial striatum in instrumental conditioning. Cereb. Cortex *20*, 873–883.

Li, C., and Rainnie, D.G. (2014). Bidirectional Regulation of Synaptic Plasticity in the Basolateral Amygdala Induced by D1-like Family of Dopamine Receptors and Group II Metabotropic Glutamate Receptors. J. Physiol. *19*, 4329–4351.

Maltais, S., Côté, S., Drolet, G., and Falardeau, P. (2000). Cellular colocalization of dopamine D1 mRNA and D2 receptor in rat brain using a D2 dopamine receptor specific polyclonal antibody. Prog. Neuro-Psychopharmacology Biol. Psychiatry *24*, 1127–1149.

Marowsky, A., Yanagawa, Y., Obata, K., and Vogt, K.E. (2005). A specialized subclass of interneurons mediates dopaminergic facilitation of amygdala function. Neuron *48*, 1025–1037.

Mashhoon, Y., Tsikitas, L.A., and Kantak, K.M. (2009). Dissociable effects of cocaine-seeking behavior following D1 receptor activation and blockade within the caudal and rostral basolateral amygdala in rats. Eur. J. Neurosci. *29*, 1641–1653.

Muller, J.F., Mascagni, F., and McDonald, A.J. (2009). Dopaminergic innervation of pyramidal cells in the rat basolateral amygdala. Brain Struct. Funct. *213*, 275–288.

Muly, E.C., Maddox, M., and Khan, Z.U. (2010). Distribution of D1 and D5 dopamine receptors in the primate nucleus accumbens. Neuroscience *169*, 1557–1566.

Namburi, P., Beyeler, A., Yorozu, S., Calhoon, G.G., Halbert, S.A., Wichmann, R., Holden, S.S., Mertens, K.L., Anahtar, M., Felix-Ortiz, A.C., et al. (2015). A circuit mechanism for differentiating positive and negative associations. Nature *520*, 675–678.

Sah, P., Faber, E.S.L., Lopez De Armentia, M., and Power, J. (2003). The amygdaloid complex: anatomy and physiology. Physiol. Rev. *83*, 803–834.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. J. Neurophysiol. *80*, 1–27.

Simmons, D.A., Brooks, B.M., and Neill, D.B. (2007). GABAergic inactivation of basolateral amygdala alters behavioral processes other than primary reward of ventral tegmental self-stimulation. Behav. Brain Res. *181*, 110–117.

Swanson, L.W. (1982). The projections of the ventral tegmental area and adjacent regions: A combined fluorescent retrograde tracer and immunofluorescence study in the rat. Brain Res. Bull. *9*, 321–353.

Tye, K.M., Prakash, R., Kim, S.-Y., Fenno, L.E., Grosenick, L., Zarabi, H., Thompson, K.R., Gradinaru, V., Ramakrishnan, C., and Deisseroth, K. (2011). Amygdala circuitry mediating reversible and bidirectional control of anxiety. Nature *471*, 358–362.

Wall, N.R., De La Parra, M., Callaway, E.M., and Kreitzer, A.C. (2013). Differential innervation of direct- and indirect-pathway striatal projection neurons. Neuron *79*, 347–360.

Wallace, D.M., Magnuson, D.J., and Gray, T.S. (1992). Organization of amygdaloid projections to brainstem dopaminergic, noradrenergic, and adrenergic cell groups in the rat.

Brain Res. Bull. *28*, 447–454.

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. Eur. J. Neurosci. *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. Eur. J. Neurosci. *22*, 513–523.

Zahm, D.S., Jensen, S.L., Williams, E.S., and Martin III, J.R. (1999). Direct comparison of projections from the central amygdaloid region and nucleus accumbens shell. Eur. J. Neurosci. *11*, 1119–1126.

"A story has no beginning or end: arbitrarily one chooses that moment of experience from which to look back or from which to look ahead."

Graham Greene

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt