CURRICULUM VITAE

2008 - 2011 BSc Medicine
VU University Amsterdam

2012 - 2014 MSc Neuroscience
(cum laude)
VU University Amsterdam

2014 - 2018 PhD
Utrecht University
Adan & Vanderschuren labs

2018 - now Postdoc
UC Berkeley
Lammel lab

linkedin.com/in/jeroenverharen

Scholar: Jeroen P.H. Verharen

ABOUT THE COVER

Traces on the cover represent a population response of rat dopamine neurons to the receipt of reward (blue) and punishment (red); data from chapter 2.

NEUROECONOMIC MECHANISMS OF REWARD AND AVERSION

JEROEN P.H. VERHAREN

NEUROECONOMIC MECHANISMS OF REWARD AND AVERSION

JEROEN P.H. VERHAREN

Brain Center Rudolf Magnus

Universiteit Utrecht

ISBN 978-90-393-7059-9

253

Brain Center Rudolf Magnus

UITNODIGING

Voor het bijwonen van de openbare verdediging van het proefschrift

NEUROECONOMIC MECHANISMS OF REWARD AND AVERSION

op dinsdag 8 januari 2019 om 12:45 uur in het Academiegebouw van de Universiteit Utrecht, Domplein 29 in Utrecht.

Aansluitend volgt een receptie.

CONTACT

Jeroen Verharen
jeroenverharen@gmail.com

Paranimfen:
Tara Arbab
t.arbab@nin.knaw.nl
Kim Mertens
mertenskl@gmail.com

NEUROECONOMIC MECHANISMS OF REWARD AND AVERSION

JEROEN P.H. VERHAREN

Neuroeconomic mechanisms of reward and aversion

Neuroeconomische mechanismen van beloning en straf

(met een samenvatting in het Nederlands)

Colophon

Author	Jeroen Verharen
Cover	Jeroen Verharen & Jeroen Derks
Layout	Jeroen Verharen
Press	Gildeprint - www.gildeprint.nl
ISBN	978-90-393-7059-9

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties in het openbaar
te verdedigen op dinsdag 8 januari 2019 des middags te 12.45 uur

door Jeroen Petrus Hendrikus Verharen

geboren op 12 juli 1989 te Rossum

Promotoren: Prof.dr. R.A.H. Adan
Prof.dr. L.J.M.J. Vanderschuren

Acknowledgements

To Roger and Louk, for their guidance and lessons.
To Mieneke, for the many hours of surgeries and perfusions.
To the animal caretakers, technicians and staff of BCRM, for all their help.
To all the Adan & Vanderschuren lab members, for the discussions.

Aan mijn vrienden, voor de afleiding.
Aan mijn familie, voor de steun.
En aan Dennis, voor alles.

“Whenever choice appears in any form - as rivalry between appetites which cannot be simultaneously satisfied, as a perceived meaning attached to an ambiguous stimulus, as a planned decision between two courses of action, as a symbolic fulfillment of an unsuspected act - it always involves an element of inhibition.”

Solomon Diamond, Richard Balvin and Florence Rand Diamond, in *Inhibition and choice: A neurobehavioral approach to problems of plasticity in behavior* (1963)

CHAPTER 1 10
Introduction

CHAPTER 2 30
A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states

CHAPTER 3 74
Regional specialization of value-based learning functions in the rat prefrontal cortex

CHAPTER 4 92
Differential contributions of striatal dopamine D1 and D2 receptors to value-based learning and decision making

CHAPTER 5 112
Reinforcement learning across the rat estrous cycle

CHAPTER 6 124
Corticolimbic mechanisms of behavioral inhibition under threat of punishment

CHAPTER 7 144
Dopaminergic contributions to behavioral control under threat of punishment in rats

CHAPTER 8 160
Limbic control over the homeostatic need for sodium

CHAPTER 9 178
Insensitivity to monetary losses in anorexia nervosa patients

CHAPTER 10 194
General discussion

CHAPTER 11 204
Addendum

TABLE OF CONTENTS

CHAPTER 1

Introduction

Neuroeconomic mechanisms of reward and aversion

CHAPTER 1

Reward and aversion

In order to survive and flourish in a competitive world, an organism must learn to repeat actions that have proven profitable and avoid actions that have not. In this way, one learns to adapt its behavior in a changeable environment, in order to optimally promote survival. For example, it is smart to revisit a place that is rich in foods, but not when this same place is filled with predators. By incorporating these rewarding (food) and negative (predator) experiences into a value representation of stimuli in the surrounding world, one can enjoy rewards, such as food and sex, without experiencing life-threatening dangers. These value representations, and the repeated updating of these values based on each action's outcome, are the essence of decision-making processes that living organisms encounter numerous times each day.

Adapting behavior in response to positive and negative feedback is driven by a learning process called operant conditioning or instrumental learning. First stated by Thorndike[1], and later refined by Skinner[2,3] (see also Box 1), is the notion that cats, pigeons and rats tend to increase the frequency of a certain behavior when this behavior is reinforced - either by the delivery of something pleasant (positive reinforcement) or the removal of something aversive (negative reinforcement). Thorndike described this theory in his *Law of effect*[1], after observing that a cat that is restrained in a box gradually learns how to escape by using trial and error. 40 years after Thorndike's cat experiments, Skinner set the stage for the next decades of experimental psychological research by theorizing operant conditioning in his book *The behavior of organisms*[2] and the development of the now widely used operant conditioning chambers. Although his theory was more formally postulated than Thorndike's, the idea behind the theory remained the same: behavior that is reinforced will be repeated, and behavior that is punished will cease. The operant conditioning chambers that he created became the world standard laboratory tool to study how reward and aversion shape behavior of animals, and these are still a key tool for animal research on addiction, decision making, and learning and memory.

In the last decades, interest in operant conditioning has increased due to the rise of artificial intelligence and its subfield of machine learning, which studies the ability of computers to learn on the basis of data without being explicitly programmed. One form of machine learning is called reinforcement learning, which teaches computers how to ideally respond on the basis of feedback, and is essentially a quantitative approach to operant learning. As such, the computer uses positive and negative feedback to improve its own performance. Since its development, reinforcement learning has been applied to a wide variety of concepts, including computer-driven stock trading[4], teaching a computer how to play video games[5], and teaching robots how to move around in an environment[6].

A paper that is considered the foundation of reinforcement learning theory is work published by Rescorla and Wagner in 1972[7], who for the first time provided a quantitative explanation for conditioning by showing that 'surprise', i.e., a difference between anticipated and actually received reward, is the driving force behind learning. This theory was later extended by Sutton and Barto[8,9] to learning from rewards that are temporally separated from its predictive cue or preceding action. The essence of this behavioral approach to reinforcement learning is that an organism makes decisions in order to maximize reward in the long term. For example, if a hungry rat performs a behavioral task in an operant cage, it tries to earn as many rewards (e.g., food pellets) as possible.

In humans, decision making behavior entails a complex process in which the gains and costs associated with different courses of action at any particular moment in time are compared in order to maximize reward. Such a reward can be anything, from the consumption of a delicious snack to maximizing profits during a night in the casino, to going to college in order to achieve long-term wealth and safety. As in other organisms, reinforcement learning plays a mediating role in these decision-making processes; for each

possible action, one makes a cost/benefit analysis on the basis of previous experiences. These costs and benefits are adjusted for its probability of occurrence and expected timing of the outcomes. For example, when you want to buy a tasty dessert, you consider the direct reward associated with the consumption, and penalize this in some way for the direct financial costs of the purchase, as well as the long-term health consequences of the dessert. As such, for every decision you make, the pros and cons will be weighed into a net expected value which will determine whether you will perform a certain action or not (which logically often results in doing nothing).

Neuronal value signals

Given the large number of decisions an organism has to make on a daily basis, one would expect that a large part of our brain is dedicated to value coding, feedback integration, and value comparisons. In the past decades, many of such value-related brain signals have been observed using various neuroimaging and neuronal recording techniques. A formal distinction can be made between a reward signal, in which neuronal activity changes during reward delivery, and a reward prediction error signal, in which neuronal activity changes in response to the 'surprise' associated with unexpectedly occurring reward or rewarding stimuli.

A value signal is a special case of a reward signal that scales with the subjective intensity of the reward. This intensity can reflect both differences in quantity (a bigger reward will yield a higher neuronal response) and quality (a better reward will yield a higher neuronal response). Furthermore, these value signals could, in principle, represent a net expectation, i.e., the expected value associated with a certain action after subtraction of its costs (i.e., effort and aversive consequences). One could logically assume that in order to make such computations, there must be some sort of common currency, i.e., a single "one size fits all" scale of value, that can be used to compare choice options of different modalities (for example, choosing between coffee and a banana).

Evidence in favor of neuronal value coding in such a common currency comes from a landmark study by Padoa-Schioppa & Assad (2006)[10], who performed single unit recordings in the orbitofrontal cortex of rhesus monkeys. Animals could choose between two types of juices that differed in taste and were offered in different quantities on a visual screen, and the monkeys could make a choice between a left and right offer by making eye movements. They found that during the choice process, the majority of neurons in the orbitofrontal cortex encoded some aspect of the choices the monkeys made. These neurons either encoded 1) the quantity of one of the offered juices, 2) the value (a combination of taste and amount) of the chosen juice, or 3) the taste of the chosen juice (a binary response to one of the two juices during reward delivery). In a follow-up study, these authors demonstrated that the neuronal responses to an offered or chosen reward did not depend on which other rewards were offered at the same time[11]. Collectively, these data point towards orbitofrontal cortex neurons encoding aspects of choice in a single, common value measure that can compare qualitatively different options. A recent study showed that during deliberation of a binary choice, these orbitofrontal cortex neuronal ensembles that encode the two different option values alternate in activity, suggesting that these neurons are directly involved in weighing choice options[12]. Similar forms of economic value coding have later been found in the ventromedial region of the prefrontal cortex of monkeys[13]. Despite various efforts, no direct evidence has thus far been found that individual brain cells of rodents encode value in a single, common scale.

Whether these neuronal value signals are subsequently compared and courses of actions selected by distinct, downstream brain regions remains a matter of debate[14,15]. In contrast to the modular view on economic choice, in which each brain region controls one chain of the choice process, some researchers have proposed that during decision making,

multiple brain regions compute value components of choice independent of each other[16-18]. A parallel is drawn with the distributed decision making of bee swarms: when looking for a potential new hive site, the bees make a choice for a new site in concert, through a distributed consensus, emerging from the information gathered by individual bees[19]. Likewise, different brain areas evaluate, compare and/or select different choice options, and a choice emerges as a result of the interactions of these regions on a circuit level[16,18]. One paper[20] suggested that different brain regions have a role in disentangling the different aspects of choice from its sensory information, very similar to how the visual system delineates visual imageries. As a result, brain regions involved in decision making encode abstract decision making variables that each retain components of the value of the options and thus physiologically demonstrate value correlates. This may explain why reward signals have been observed throughout the brain, and suggests that there is no final common pathway for choice selection, but that value signals converge at multiple points to eventually compete for execution in the motor system.

There is substantial evidence that aversive stimuli are also explicitly coded in the brain. For example, lateral habenula neurons have shown to increase activity in response to unexpected punishment and decrease activity in response to unexpected reward[21,22]. Furthermore, a subpopulation of basolateral amygdala neurons projecting to the central amygdala are primarily activated by aversive stimuli, and these have been shown to be essential for fear conditioning[23,24]. The regions implicated in processing aversive stimuli are partially overlapping with the regions involved in processing rewarding stimuli, and include the nucleus accumbens, septum, prefrontal cortex, amygdala and hippocampus[25].

Reward prediction error signals

During value-based learning, expectations of reward after actions and stimuli (regardless whether this is encoded in a common scale of value or not, and whether this is processed in series or in parallel) are updated on the basis of experiences, creating an up-to-date representation of reward value of the surrounding world that is necessary for making profitable decisions. As postulated by reinforcement learning theories, this updating process may be guided by prediction errors, or 'surprise', computed by subtracting the received reward from the cached reward expectation:

Reward prediction error = Reward received - Reward expected (1)

As such, when a reward is better than expected (i.e., a positive reward prediction error), the value associated with the action or stimulus that preceded that reward should be increased, and when a reward is worse than expected or when explicit punishment has occurred (i.e., a negative reward prediction error), the value of the preceding action or stimulus should be decreased.

Thus, a reward that is fully predicted by a preceding sensory stimulus will not evoke a neuronal response during the reward itself, as the surprise (i.e., reward prediction error) associated with that reward will be 0. Neurons that encode reward prediction errors will therefore, after extensive learning, only show changes in activity during the conditioned stimulus that precedes the reward or punishment, but not the unconditioned stimulus itself (Figure 1). Conversely, when an expected reward is not delivered, or when explicit punishment is delivered, a negative reward prediction error occurs, resulting in a reduction in firing rate. Such positive and negative reward prediction errors are thought to be important mediators of approach and avoidance learning, respectively[26-28].

Although theoretically and physiologically distinct, it can be quite challenging to discern experimentally between reward signals, reward prediction error signals, and for example general responses to salient stimuli (Figure 1). To have a full transfer of the

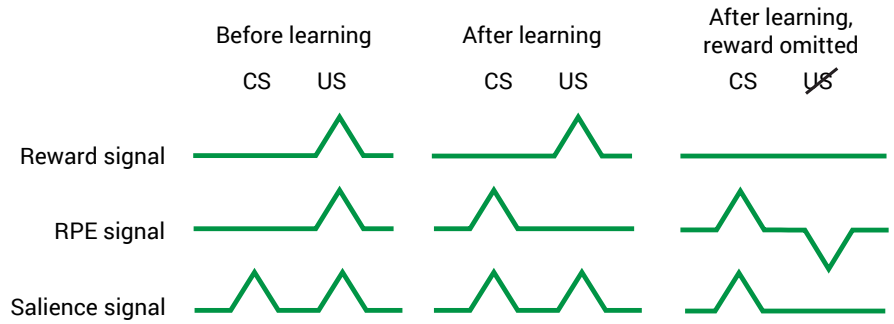


Figure 1: Reward and reward prediction error (RPE) signals in the brain. After extensive training, reward prediction errors signals will only emerge during the conditioned (CS; predictive cue), but not unconditioned (US; reward) stimulus.

neuronal signal from the unconditioned to the conditioned stimulus, animals need to have fully learned the association (which may require a long training period), the environment should be perfectly predictable, and timing of the occurrence of the unconditioned stimulus by the experimental subject should be precise. Many studies report neuronal activation during both the conditioned and unconditioned stimuli (e.g., refs. 24,29,30), suggesting that these requirements have not been met or that mixed neuronal signals have been recorded.

The role of dopamine

Although neuronal signals with characteristics of reward prediction error have been found across a wide range of brain areas[27,31], the neurocomputationally most pure and perhaps behaviorally most essential form of prediction error coding is found in dopamine cells in the midbrain[31,32]. A large proportion of these neurons have been shown to increase firing in response to better-than-expected reward, to decrease firing in response to worse-than-expected reward or explicit punishment, and to show no change in firing when reward is fully predictable — an observation that has been reported in a wide range of species including humans[33], monkeys[34,35] and rodents[32,36]. In the last decades, dopamine neurons have therefore emerged as a prime candidate in mediating reinforcement learning.

The first major evidence for an involvement of dopamine in reward processing came from influential work in 1954 from Olds and Milner that showed that animals vigorously lever press in exchange for electrical stimulation of limbic brain structures[37] (Figure 2), a phenomenon now known as intracranial self-stimulation. Although this first experiment was not performed directly in the dopamine system, follow-up studies showed that intracranial self-stimulation was strongest for midbrain dopamine nuclei and connected regions, and that half of all the brain regions for which animals showed self-stimulation were directly connected to dopamine neurons[38-41]. A role for dopamine in mediating reinforcement was further suggested by a series of studies that showed that operant responding for natural rewards and addictive drugs could be attenuated by pharmacological blockade of dopamine receptors in the brain[42-44].

Interest in dopamine sparked when Schultz and colleagues[26] made an exciting discovery in the 90's: they found neuronal correlates of reward prediction errors in midbrain dopamine neurons of monkeys, exactly as predicted by Rescorla and Wagner[7] more than two decades earlier, and in accordance with Sutton and Barto's temporal difference learning model[8,9]. This discovery was an important step in the understanding of dopamine function

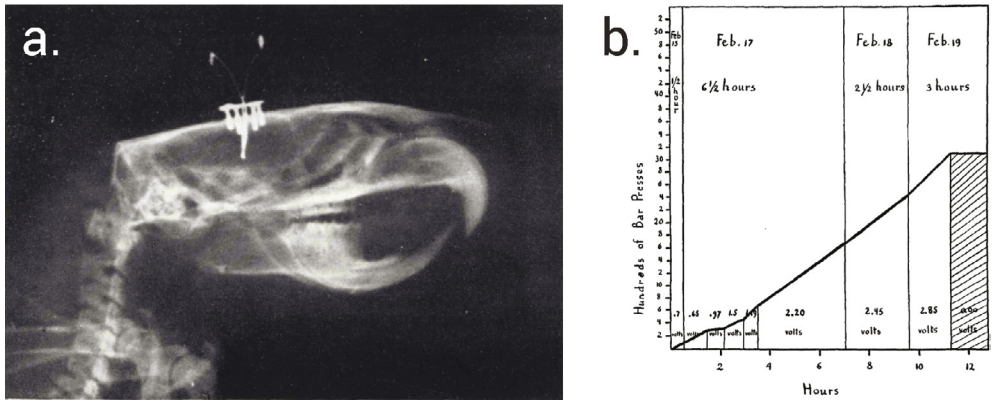


Figure 2: Images from the original Olds & Milner paper who for the first time demonstrated that animals will lever press for electrical stimulation of limbic brain structures. a) X-ray image of a rat with an electrode implant. b) Learning curve of an animal implanted with an electrode in the septal area making lever presses for electrode stimulation.

and suggested a direct role for dopamine neurons in reward and punishment learning, thereby mediating motivation and decision making[28,41,45].

Although the importance of dopaminergic prediction errors to learning was quickly acknowledged, their sufficiency for learning has been confirmed only recently, employing optogenetic tools in rodents. In their study, Steinberg et al. (2014) demonstrated that brief optogenetic activation of VTA dopamine neurons was able to drive learning of the association between a conditioned stimulus and reward[46]. They further showed that activation of dopamine neurons during the time of expected reward delivery slowed extinction learning, together suggesting that artificial positive reward prediction error is sufficient to drive appetitive learning. Conversely, Chang et al. (2015) showed that brief optogenetic inhibition of VTA dopamine neurons of mice was sufficient to mimic negative reward prediction error and thereby drive avoidance learning[47]. Finally, Saunders et al. (2018) demonstrated that optogenetic excitation of VTA dopamine neurons during a cue was sufficient to attribute incentive motivational value to that cue, even in the absence of natural reward, and showed that this was driven by dopaminergic neurotransmission in the core region of the nucleus accumbens[48].

To compute reward prediction error, a system needs, by definition, information about the reward it expects. Takahashi et al. (2011) studied whether midbrain dopamine neurons receive this information from the orbitofrontal cortex by measuring reward prediction error in the ventral tegmental area during a reward-learning task in rats with and without a neurotoxic lesion of the lateral orbitofrontal cortex[49]. They observed that both positive and negative reward prediction error coding in the ventral tegmental area was attenuated by the lesion. However, the pattern of observed effects did not match the hypothesis that the lesioned part of orbitofrontal cortex conveyed a pure value signal to the dopamine neurons, as the authors demonstrated by simulating electrophysiological data with reinforcement learning models. More recent work suggests that dopamine neurons use value information from a wide range of areas to compute prediction errors[32]. Wherever these value signals arise from, electrophysiological evidence suggests that dopamine neurons use subtraction to compute the prediction error from the expected and received reward, and that inhibition by GABA neurons of the ventral tegmental area facilitate this computation[50].

Despite the apparent homogeneity of prediction error responses in midbrain dopamine neurons in some studies[51], it must be noted that since the development of genetic tools for neural circuit dissection, an increasing number of studies points towards heterogeneity in dopamine cells with regards to connectivity, morphology, gene expression and function[48,52-54]. Furthermore, reward-related responses of individual dopamine neurons have been shown to encode aspects of motor behavior[55,56], suggesting that prediction errors are not encoded as mathematically pure as was thought before. That said, the importance of dopamine and dopaminergic reward prediction error to value-based learning has been one of the most well-established principles in recent neuroscientific history[28,31,44,57,58].

A neuroeconomic approach to motivation

One aspect of reward-related behavior for which dopamine is critical is motivation[59,60]. Although different authors use slightly different definitions of this term[60], motivation usually refers to the willingness to work for something. This can be the receipt of reward, but also the avoidance of a punisher. As such, when dopamine in the brain is depleted, animals will cease to actively search for food (and eventually starve to death), but they will still consume food when it is placed in their mouth[61]. Berridge and Robinson therefore proposed a useful distinction between the 'liking' (i.e., experiencing pleasure) and 'wanting' (i.e., the motivation) of reward[62,63], and it has been suggested that dopamine is mainly involved in the latter[60].

In neuroeconomic terms, motivation is thought of as the subjective experience that a certain action is worth pursuing. The value of such an action can be described by an economic utility function[64], so that every time a organism considers a certain action, a computation is performed where the subjective experience of the costs (labor and negative consequences, corrected for the probability of occurrence) is subtracted from the expected reward that follows from that action (receiving food, sex, drugs or shelter, or avoiding punishment, corrected for probability)[65], yielding the net expected reward value associated with that action:

$$\text{Net expected reward} = \sum \text{reward}_{\text{subjective}} - \sum \text{costs}_{\text{subjective}} \quad (2)$$

Only when this calculation has a positive outcome, an action is pursued, as the expectation of reward is higher than its cost. Conversely, when the outcome of this calculation approaches 0 or becomes negative (when costs > reward), no action is taken. The subjective reward term in this equation ($\sum \text{reward}_{\text{subjective}}$) can be seen as the expectation of pleasure associated with reward ('liking'), and the outcome of the equation is proportional to motivation ('wanting'), so that:

$$\text{Motivation} \propto \sum \text{reward}_{\text{subjective}} - \sum \text{costs}_{\text{subjective}} \quad (3)$$

For example, whether an animal will start foraging for food depends on several factors. First, it depends on the amount of food it expects to receive, which is the total food that is expected to be obtained from the environment ($\sum$ reward). Second, it depends on to what extent the food is appreciated; a satiated animal will appreciate food less than a hungry animal. Hence, the objective reward expectation $\sum$ reward should be multiplied with a subjectivity factor that reflects the metabolic and hedonic state of the animal, leading to a subjective reward value $\sum \text{reward}_{\text{subjective}}$. Conversely, the costs of foraging depends on the effort the animal has to make to seek for food and the dangers associated with food seeking (i.e., the probability of explicit negative consequences, like a predator attack). Again, this factor should be corrected for subjectivity, leading to a subjective cost factor $\sum \text{costs}_{\text{subjective}}$. When the expected reward outweighs the costs, the animal will start foraging. Logically, subjective reward value increases with hunger (a meal tastes much better when you're

hungry), so that even in a dangerous environment, reward will at some point outweigh costs, and the motivation to start to seek for food will increase. Furthermore, influential economic and psychological theories state that rewards and costs that are further in the future or that are less likely to be received are discounted, i.e., its subjective value is reduced with time and probability — a process known as temporal or probability discounting, respectively[66-69]. Hence, equation 3 can be rewritten as:

$$\text{Motivation} \propto \sum s_{reward} * \gamma_{reward} * \text{reward} - \sum s_{costs} * \gamma_{costs} * \text{costs} \quad (4)$$

in which *s* represents a subjectivity factor that scales the reward/cost on the basis of the animal's intrinsic state and desires, and *γ* a discounting factor that is low when the rewards or costs are further away in the future or are less likely to occur.

This simple framework of motivation may help structuring our understanding of phenomena that are associated with reward seeking and motivation. For example, the vast increase in the prevalence of obesity in the Western world[70] is thought to arise from the difficulty to make healthy food choices and the fact that it is hard to lose weight[71]. In our society, the costs associated with food intake are radically different than they have been for the past millennia and different than for animals in the wild. For animals and premodern man, the costs mainly comprised the physical effort and the dangers that were associated with hunting and other forms of foraging. For modern man, given the abundance of food, the costs comprise the financial costs of the food and the negative health consequences that are associated with food intake. Given that food is usually directly available, equation 4 can be given by:

$$\text{Motivation} \propto s * [\text{food reward}] - s * \gamma * [\text{health consequences}] - s * [\text{financial costs}] \quad (5)$$

Despite the potential severity of the health consequences of palatable foods, they often develop over a longer period of time and are thus not immediately noticed. This discounts the subjective experience of the negative health consequences to a negligible level, except perhaps when someone has low temporal discounting characteristics. Indeed, trait impulse control, closely associated to temporal discounting capacity, is predictive for the maintenance of overweight in children[72] and adults[73]. An additional point is that unhealthy foods, high in carbohydrates and fat, are usually cheaper than healthy foods, adding an extra costs factor to the equation, thereby decreasing the motivation to make healthy food choices — a factor that may especially play a role in people with a low income[74]. Thus, the direct reward of palatable food and the absence of any direct costs associated with its intake makes it ostensibly unprofitable to make healthy food choices. Limiting palatable food intake is especially hard during dieting, as this in fact increases sensitivity to food reward[75,76], making the left side of this equation more dominant.

A second useful application of this framework is to understand drug addiction and the fact that some people are more prone to develop this mental disorder than others. Every time a drug user gets reminded of drugs (by cravings, cues or social pressure), he or she will make a decision to take drugs or not. Considering the expectation of reward from the 'high' of the drug and the negative consequences of drug use (financial costs, hangovers, long-term health consequences and consequences for social obligations), equation 4 can be written as:

$$\begin{aligned}\text{Motivation} \propto\ & s * [\text{high}] \\ & - s * \gamma * [\text{financial costs}] \\ & - s * \gamma * [\text{hangover}] \\ & - s * \gamma * [\text{health consequences}] \\ & - s * \gamma * [\text{social consequences}] \end{aligned} \quad (6)$$

Given this list of costs associated with prolonged drug use, it is not surprising that only a small fraction of recreational drug users eventually develops addiction[77]. Based on this equation, however, several risk factors can be derived for the development of addiction. First, increased expectation sensitivity to the 'high' of the drugs (note that this is different from the actually received pleasure from the drugs), which would strengthen the left side of equation 6. Second, low baseline levels of the costs factors — that is, a poor social life, no job or study, and bad health —, making the costs of drug use low. Third, a low value of temporal discounting factor γ (i.e., discounting of subjective value over time is stronger). Indeed, clinical and preclinical studies have demonstrated that all of these factors increase the risk for the development of addiction[77-80]. Additionally, negative consequences of repeated drug use directly decrease baseline levels of health and social life, essentially decreasing the cost factors in this equation, thus making future drug taking more tempting. Furthermore, cocaine administration itself chronically strengthens temporal discounting capability[81,82], indicating that dopaminergic drugs directly increase the motivation to take drugs.

Implications for psychiatry

Abnormalities in the brain circuits involved in value processing, motivation and decision making have been implicated in a wide variety of neuropsychiatric disorders. For example, dysfunctions in the dopamine system have been implicated in depression[83-85], addiction[44,86], bipolar disorder[87,88], ADHD[89,90] and schizophrenia[91] — not the least because most of the effective pharmacotherapies for some of these diseases target the dopamine system. Moreover, dysfunction of the prefrontal cortex, another important region for value-based decision making, has been implicated in a partially overlapping set of disorders, including addiction, impulse control disorders, depression, schizophrenia, autism and ADHD. Besides dysfunctions in these brain circuits, altered value-based decision making has been observed in all of these patient groups[92-101], an indication that alterations in value processes might be involved in the etiology of these diseases. Whether this relationship is causal and driven by miscalculations on a neuronal level remains a challenging question in neuroscience, although remarkable progress has been made in this regard.

For example, it has been suggested that depression at least partially arises from unrealistically low reward expectations, mainly due to pessimistically set priors (i.e., assumptions), although it has been shown that abberations in model-free learning mechanisms, as assessed by classical reinforcement learning models, are likely not involved in the pathophysiology of the disease[102]. Furthermore, neurocomputational models predicted that the reckless and overoptimistic decision-making behavior after levodopa treatment in Parkinson's disease patients is induced by impaired prediction error learning due to overstimulation of striatal dopamine receptors[103,104]. This hypothesis has been supported by several clinical studies[105-107] and may be of potential impact on the understanding of mania, as this mental state is also associated with elevated dopamine levels[96,108]. A third example is anxiety, which has been suggested to result from increased threat avoidance due to an overestimation of the probability and magnitude of aversive outcomes, a mechanism that may arise from alterations in brain areas involved in learning and value-based decision making, like the amygdala and anterior cingulate cortex[109].

The recent emergence of several new methods for computational analyses, large-scale neuronal recordings and neuronal manipulations now allow for a precise investigation of the neural circuits involved in reward and punishment processing. Based on recent findings employing these techniques, it is obvious that the neuronal circuits involved are more heterogeneous than previously thought, suggesting that we have only just begun to elucidate the computational basis of neuropsychiatric disorders.

Box 1

A brief history of research on reward, aversion and motivation

1848 **Harlow** publishes the case report on Phineas Gage, providing the first experimental evidence for a role of the prefrontal cortex in executive behaviors, including decision making.

1898 In his *Law of effect*, **Thorndike** states that animals learn through trial and error, an important step in the postulation of operant conditioning theory.

1927 **Pavlov** formulates his associative learning theory on the basis of his famous dog experiment.

1938 **Skinner** publishes *The behavior of organisms*, including the influential theory on operant conditioning.

1946 **Tolman** challenges earlier conditioning theories by stating that learning can also occur in the absence of reward or punishment (i.e., stimulus-stimulus learning).

1954 **Olds and Milner** discover that rats will work for electrical stimulation of certain brain areas, a phenomenon now known as intracranial self-stimulation.

1972 Publication of the influential reinforcement learning theory of **Rescorla and Wagner**, proposing that prediction errors drive learning.

1982 **Dickinson** performs a set of experiments in rats that demonstrate a distinction between goal-directed and habitual behavior.

1981 **Sutton and Barto** publish computational models that explain temporal difference learning.

1997 The first measurement of reward prediction error signals in dopamine neurons of monkeys by **Schultz**.

1998 **Berridge and Robinson** propose their incentive salience theory of dopamine function, introducing the dichotomy between 'wanting' and 'liking'.

2007 **Boyden, Deisseroth, Roth** and others develop viral tools to manipulate brain activity: start of the era of neural circuit dissection.

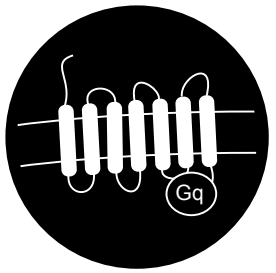
Outline of this thesis

The aim of this thesis is to gain insight into the neurocomputational basis of decision-making, and specifically how learning from reward and aversion shapes choice behavior. To achieve this, we combined several decision-making paradigms with chemogenetics, behavioral pharmacology, fiber photometry and computational methods.

Chapters 2 through 5 mainly focus on the learning component of value-based decision making. In **chapter 2**, we study the neural basis of the aberrations in choice behavior that are associated with an abundance of dopamine in the brain, for example during mania, the 'high' of drugs of abuse, and dopamine replacement therapy in Parkinson's disease. We specifically test a theory based on a neurocomputational model of striatal function that implicates impaired negative reward prediction error processing in these decision making deficits. In **chapter 3**, we study the contribution of different regions of the prefrontal cortex to value-based learning, by pharmacological inactivating these regions in rats during a value-based learning task. Employing this same task, in **chapter 4** we describe how treatment with dopamine receptor agonists and antagonists affects value-based learning, thereby trying to pinpoint receptor-specific contributions to the negative feedback learning effect that we describe in chapter 2. To demonstrate that value-based learning is a dynamic process that is dependent on the state of an organism, we show in **chapter 5** that computational processes that underlie this behavior fluctuate across the estrous cycle of female rats.

Chapters 6 through 9 focus more on the motivational aspects of value-based decision making, with a special focus on maladaptations in the systems that regulate food intake. In **chapter 6**, we describe behavior in a task that can be seen as a form of irrational decision-making, namely loss of control over behavior when a choice has to be made between pursuing reward and avoiding punishment. We show proof-of-concept by combining the task with pharmacological inactivations of different regions of the corticolimbic system, and show that inhibition of reward pursuit requires the coordinated action of a network of structures in this system. We further utilize this behavioral task in **chapter 7**, where we combine it with different viral and pharmacological techniques to elucidate the role of the dopamine system in behavioral control. In **chapter 8**, we study salt appetite, and the extent to which midbrain dopamine neurons mediate this process. Salt appetite refers to the fact that salty solutions are normally considered aversive, but suddenly become appetitive when the body is in a sodium-depleted state. This switch in salt appreciation may teach us a lot about the flexibility of brain reward systems. Furthermore, we present an experimental human study in **chapter 9**, in which we investigate decision making in anorexia nervosa patients, by means of computational modeling of a dataset of a large cohort of patients that performed the Iowa Gambling task. Finally, I will discuss the findings from this thesis and try to place it into existing literature in the **Discussion**.

Techniques used in this thesis



Chemogenetics

Chemogenetics or DREADD (Designer Receptor Exclusively Activated by Designer Drugs) is a viral tool used to manipulate brain activity in a cell type- or projection-specific manner. DREADDs are mutated receptors that, after activation by its ligand clozapine-N-oxide (CNO), can depolarize (activation; Gq DREADD) or hyperpolarize (inhibition; Gi DREADD) a cell. The genes for DREADD expression are delivered by an intracranial injection of an adeno-associated virus.

Used in chapters 2 and 7.

Behavioral pharmacology

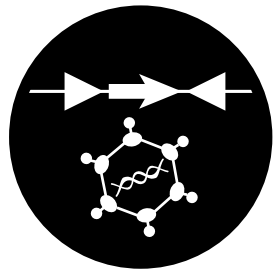
Behavioral pharmacology refers to the use of biologically active agents to study the contribution of certain brain areas, receptors and cell types to behavior. Typically, receptor agonists and antagonists are injected systemically or infused through cannulas that are implanted above a brain area of interest.

Used in chapters 2, 3, 4, 6, 7 and 8.

Fiber photometry

Fiber photometry is a method to measure neuronal activity in living animals using calcium-dependent fluorescent proteins (genetically encoded calcium indicators; GCaMPs). After viral delivery of these GCaMPs, a fiber is implanted above the transfected cell bodies, capturing the bulk fluorescence, which can be used as a measure for neuronal population activity.

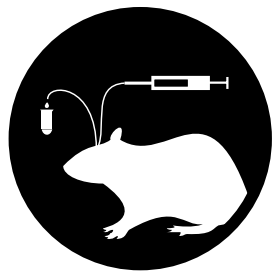
Used in chapters 2, 7 and 8.



Cre-lox

The Cre-lox system is a genetic tool to achieve cell-type or projection specificity in gene expression. If a Cre-dependent viral vector is used, the gene of interest will only be expressed in cells that express the protein Cre. For example, by injecting a Cre-dependent virus in tyrosine hydroxylase (TH)::Cre animals, a viral construct will only by expressed in TH-positive cells (i.e., cells that produce the neurotransmitters dopamine and noradrenaline).

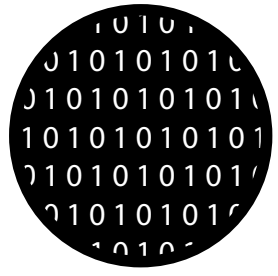
Used in chapters 2, 7 and 8.



Microdialysis

Microdialysis can be used to measure the extracellular concentrations of different types of molecules (such as dopamine). A semi-permeable probe is placed into the brain through which a fluid is perfused. Molecules present in the extracellular space will diffuse into the liquid inside this probe, and this liquid can be collected and analysed using high-performance liquid chromatography.

Used in chapter 2.



Computational modeling

Computational modeling is an umbrella term used to indicate that an algorithm was used to find patterns in complex data.

Used in chapters 2, 3, 4, 5 and 9.

cFos immunoreactivity

cFos is an immediate early gene that is expressed in most neurons after depolarization. cFos protein levels peak ~90 minutes after neuronal activation and can be visualized using immunohistochemistry to get a readout of brain activity.

Used in chapters 7 and 8.

References

1. Thorndike, E. L. Animal Intelligence: An Experimental Study of the Associative Processes in Animals. Psychological Review 5, 551-553 (1898).
2. Skinner, B. F. The Behavior of Organisms: An Experimental Analysis. (Appleton Century Crofts, 1938).
3. Skinner, B. F. Are theories of learning necessary? Psycholog. Rev. 58, 193-216 (1950).
4. Jae Won, L. in ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No.01TH8570). 690-695 vol.691.
5. Mnih, V. et al. Human-level control through deep reinforcement learning. Nature 518, 529 (2015).
6. Peters, J., Vijayakumar, S. & Schaal, S. in Humanoids2003, Third IEEE-RAS International Conference on Humanoid Robots (Karlsruhe, Germany, 2003).
7. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 2, 64-99 (1972).
8. Sutton, R. S. & Barto, A. G. Towards a Modern Theory of Adaptive Networks: Expectation and Prediction. Psychological Review 88, 135-170 (1981).
9. Sutton, R. S. & Barto, A. G. Reinforcement learning: An introduction. (MIT press, 1998).
10. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. Nature 441, 223-226 (2006).
11. Padoa-Schioppa, C. & Assad, J. A. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nature neuroscience 11, 95-102 (2008).
12. Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. Nature neuroscience 19, 973-980 (2016).
13. Strait, C. E., Blanchard, T. C. & Hayden, B. Y. Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. Neuron 82, 1357-1366 (2014).
14. Vlaev, I., Chater, N., Stewart, N. & Brown, G. D. A. Does the brain calculate value? Trends in cognitive sciences 15, 546-554 (2011).
15. Fumagalli, R. The futile search for true utility. Economics & Philosophy 29, 325-347 (2013).

16. Cisek, P. Making decisions through a distributed consensus. Current opinion in neurobiology 22, 927-936 (2012).
17. Rushworth, M. F., Kolling, N., Sallet, J. & Mars, R. B. Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Current opinion in neurobiology 22, 946-955 (2012).
18. Hunt, L. T. & Hayden, B. Y. A distributed, hierarchical and recurrent framework for reward-based choice. Nature Reviews Neuroscience 18, 172 (2017).
19. Seeley, T. D., Visscher, P. K. & Passino, K. M. Group Decision Making in Honey Bee Swarms: When 10,000 bees go house hunting, how do they cooperatively choose their new nesting site? American scientist 94, 220-229 (2006).
20. Yoo, S. B. M. & Hayden, B. Y. Economic Choice as an Untangling of Options into Actions. Neuron 99, 434-447 (2018).
21. Matsumoto, M. & Hikosaka, O. Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447, 1111-1115 (2007).
22. Bromberg-Martin, E. S. & Hikosaka, O. Lateral habenula neurons signal errors in the prediction of reward information. Nature neuroscience 14, 1209-1216 (2011).
23. Namburi, P. et al. A circuit mechanism for differentiating positive and negative associations. Nature (2015).
24. Beyeler, A. et al. Divergent Routing of Positive and Negative Information from the Amygdala during Memory Retrieval. Neuron 90, 348-361 (2016).
25. Jean-Richard-Dit-Bressel, P., Killcross, S. & McNally, G. P. Behavioral and neurobiological mechanisms of punishment: implications for psychiatric disorders. Neuropsychopharmacology 43: 1639-1650 (2018).
26. Schultz, W., Dayan, P. & Montague, P. R. A Neural Substrate of Prediction and Reward. Science 275, 1593-1601 (1997).
27. den Ouden, H. E., Kok, P. & de Lange, F. P. How prediction errors shape perception, attention, and motivation. Front Psychol 3, 548 (2012).
28. Keiflin, R. & Janak, P. H. Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. Neuron 88, 247-263 (2015).
29. Matias, S., Lottem, E., Dugue, G. P. & Mainen, Z. F. Activity patterns of serotonin neurons underlying cognitive flexibility. Elife 6 (2017).
30. Wang, D. et al. Learning shapes the aversion and reward responses of lateral habenula neurons. Elife 6 (2017).
31. Watabe-Uchida, M., Eshel, N. & Uchida, N. Neural Circuitry of Reward Prediction Error. Annu Rev Neurosci 40, 373-394 (2017).
32. Tian, J. et al. Distributed and Mixed Information in Monosynaptic Inputs to Dopamine Neurons. Neuron 91, 1374-1389 (2016).
33. D'Ardenne, K., McClure, S. M., Nystrom, L. E. & Cohen, J. D. BOLD Responses Reflecting Dopaminergic Signals in the Human Ventral Tegmental Area. Science 319, 1264-1268 (2008).
34. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. Science 275 (1997).
35. Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47, 129-141 (2005).
36. Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. Nature neuroscience 10, 1020-1028 (2007).
37. Olds, J. & Milner, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. J of Comp and Physiol Psych 47, 419-427 (1954).
38. Olds, J. Pleasure Centers in the Brain. Scientific American 195, 105-117 (1956).

39. Crow, T. J. A map of the rat mesencephalon for electrical self-stimulation. Brain Res 36, 256-273 (1972).
40. Corbett, D. & Wise, R. A. Intracranial self-stimulation in relation to the ascending dopaminergic systems of the midbrain: a moveable electrode mapping study. Brain Res 185, 1-15 (1980).
41. Schultz, W. Dopamine reward prediction error coding. Dialogues in Clinical Neuroscience 18, 23-32 (2016).
42. Gunne, L. M., Anggård, E. & Jönsson, L. E. Clinical trials with amphetamine-blocking drugs. Psychiatr Neurol Neurochir 75, 225-226 (1972).
43. Wise, R., Spindler, J., deWit, H. & Gerberg, G. Neuroleptic-induced "anhedonia" in rats: pimozide blocks reward quality of food. Science 201, 262-264 (1978).
44. Nutt, D. J., Lingford-Hughes, A., Erritzoe, D. & Stokes, P. R. The dopamine theory of addiction: 40 years of highs and lows. Nature reviews. Neuroscience 16, 305-312 (2015).
45. Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. Annual Review of Neuroscience 23 (2000).
46. Steinberg, E. E. et al. A causal link between prediction errors, dopamine neurons and learning. Nature neuroscience 16, 966-973 (2013).
47. Chang, C. Y. et al. Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nature neuroscience 19(1): 111-116 (2015).
48. Saunders, B. T., Richard, J. M., Margolis, E. B. & Janak, P. H. Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. Nature neuroscience, 21: 1072-1083 (2018).
49. Takahashi, Y. K. et al. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. Nature neuroscience 14, 1590-1597 (2011).
50. Eshel, N. et al. Arithmetic and local circuitry underlying dopamine prediction errors. Nature (2015).
51. Ungless, M. A., Magill, P. J. & Bolam, J. P. Uniform Inhibition of Dopamine Neurons in the Ventral Tegmental Area by Aversive Stimuli. Science 303, 2040-2042 (2004).
52. Lammel, S., Ion, D. I., Roeper, J. & Malenka, R. C. Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. Neuron 70, 855-862, doi:10.1016/j.neuron.2011.03.025 (2011).
53. Lammel, S., Lim, B. K. & Malenka, R. C. Reward and aversion in a heterogeneous midbrain dopamine system. Neuropharmacology 76, 351-359 (2014).
54. Morales, M. & Margolis, E. B. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. Nature reviews. Neuroscience 18, 73-85 (2017).
55. Jin, X. & Costa, R. M. Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature 466, 457-462 (2010).
56. Howe, M. W. & Dombeck, D. A. Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature 535, 505-510 (2016).
57. Hu, H. Reward and Aversion. Annu Rev Neurosci 39, 297-324 (2016).
58. Schultz, W. Dopamine reward prediction-error signalling: a two-component response. Nature reviews. Neuroscience 17, 183-195 (2016).
59. Cools, R. Role of dopamine in the motivational and cognitive control of behavior. Neuroscientist 14, 381-395 (2008).
60. Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. Neuron 76, 470-485 (2012).
61. Salamone, J. D., Correa, M., Mingote, S. M. & Weber, S. M. Nucleus accumbens dopamine and the regulation of effort in food-seeking behavior: implications for studies of natural motivation, psychiatry, and drug abuse. J Pharmacol Exp Ther 305 (2003).
62. Berridge, K. C. & Robinson, T. E. Parsing reward. Trends in Neurosciences 26, 507-513 (2003).
63. Berridge, K. C., Robinson, T. E. & Aldridge, J. W. Dissecting components of reward: 'liking', 'wanting', and learning. Curr Opin Pharmacol 9, 65-73 (2009).
64. Houthakker, H. S. Revealed preference and the utility function. Economica 17, 159-174 (1950).
65. Niv, Y., Joel, D. & Dayan, P. A normative perspective on motivation. Trends in cognitive sciences 10, 375-381 (2006).
66. Green, L., Myerson, J. & McFadden, E. Rate of temporal discounting decreases with amount of reward. Memory & cognition 25, 715-723 (1997).
67. Richards, J. B., Zhang, L., Mitchell, S. H. & de Wit, H. Delay or probability discounting in a model of impulsive behavior: effect of alcohol. J Exp Anal Behav 71 (1999).
68. Critchfield, T. S. & Kollins, S. H. Temporal discounting: Basic research and the analysis of socially important behavior. Journal of applied behavior analysis 34, 101-122 (2001).
69. Green, L. & Myerson, J. A discounting framework for choice with delayed and probabilistic rewards. Psychol Bull 130 (2004).
70. World Health Organization. Obesity: preventing and managing the global epidemic. (World Health Organization, 2000).
71. Rangel, A. & Hare, T. Neural computations associated with goal-directed choice. Current opinion in neurobiology 20, 262-270 (2010).
72. Nederkoorn, C., Braet, C., Van Eijs, Y., Tanghe, A. & Jansen, A. Why obese children cannot resist food: the role of impulsivity. Eat Behav 7, 315-322 (2006).
73. Nederkoorn, C., Houben, K., Hofmann, W., Roefs, A. & Jansen, A. Control yourself or just eat what you like? Weight gain over a year is predicted by an interactive effect of response inhibition and implicit preference for snack foods. Health Psychol 29, 389-393 (2010).
74. Steptoe, A., Pollard, T. M. & Wardle, J. Development of a measure of the motives underlying the selection of food: the food choice questionnaire. Appetite 25, 267-284 (1995).
75. Laeng, B., Berridge, K. C. & Butter, C. M. Pleasantness of a Sweet Taste during Hunger and Satiety: Effects of Gender and" Sweet Tooth". (1993).
76. van der Plasse, G. et al. Modulation of cue-induced firing of ventral tegmental area dopamine neurons by leptin and ghrelin. Int J Obes (Lond) 39, 1742-1749 (2015).
77. Jordan, C. J. & Andersen, S. L. Sensitive periods of substance abuse: Early risk for the transition to dependence. Developmental Cognitive Neuroscience 25, 29-44 (2017).
78. Kirby, K. N., Petry, N. M. & Bickel, W. K. Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. Journal of Experimental psychology: general 128, 78 (1999).
79. Kirby, K. N. & Petry, N. M. Heroin and cocaine abusers have higher discount rates for delayed rewards than alcoholics or non drug using controls. Addiction 99, 461-471 (2004).
80. Volkow, N. D. et al. Addiction: decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain's control circuit. Bioessays 32, 748-755 (2010).
81. Simon, N. W., Mendez, I. A. & Setlow, B. Cocaine exposure causes long-term

increases in impulsive choice. Behav Neurosci 121, 543-549 (2007).
82. Mendez, I. A. et al. Self-administered cocaine causes long-lasting increases in impulsive choice in a delay discounting task. Behavioral neuroscience 124, 470 (2010).
83. Nestler, E. J. & Carlezon, W. A., Jr. The mesolimbic dopamine reward circuit in depression. Biol Psychiatry 59, 1151-1159 (2006).
84. Russo, S. J. & Nestler, E. J. The brain reward circuitry in mood disorders. Nature reviews. Neuroscience 14, 609-625 (2013).
85. Han, M. H. & Nestler, E. J. Neural Substrates of Depression and Resilience. Neurotherapeutics 14, 677-686 (2017).
86. Volkow, N. D. & Morales, M. The Brain on Drugs: From Reward to Addiction. Cell 162, 712-725 (2015).
87. Berk, M. et al. Dopamine dysregulation syndrome: implications for a dopamine hypothesis of bipolar disorder. Acta Psychiatrica Scandinavica 116, 41-49 (2007).
88. Cousins, D. A., Butts, K. & Young, A. H. The role of dopamine in bipolar disorder. Bipolar disorders 11, 787-806 (2009).
89. Li, D., Sham, P. C., Owen, M. J. & He, L. Meta-analysis shows significant association between dopamine system genes and attention deficit hyperactivity disorder (ADHD). Human molecular genetics 15, 2276-2284 (2006).
90. Volkow, N. D. et al. Evaluating dopamine reward pathway in ADHD: clinical implications. Jama 302, 1084-1091 (2009).
91. Weinstein, J. J. et al. Pathway-specific dopamine abnormalities in schizophrenia. Biological psychiatry 81, 31-42 (2017).
92. Roger, R. D. et al. Dissociable Deficits in the Decision-Making Cognition of Chronic Amphetamine Abusers, Opiate Abusers, Patients with Focal Damage to Prefrontal Cortex, and Tryptophan-Depleted Normal Volunteers: Evidence for Monoaminergic Mechanisms. Neuropsychopharmacology 20, 322-339 (1999).
93. Grant, S., Contoreggi, C. & London, E. D. Drug abusers show impaired performance in a laboratory test of decision making. Neuropsychologia 38, 1180-1187 (2000).
94. Murphy, F. C. et al. Decision-making cognition in mania and depression. Psychological Medicine 31, 679-693 (2001).
95. Ernst, M. & Paulus, M. P. Neurobiology of decision making: a selective review from a neurocognitive and clinical perspective. Biological psychiatry 58, 597-604 (2005).
96. Johnson, S. L. Mania and dysregulation in goal pursuit: a review. Clin Psychol Rev 25, 241-262 (2005).
97. Shurman, B., Horan, W. P. & Nuechterlein, K. H. Schizophrenia patients demonstrate a distinctive pattern of decision-making impairment on the Iowa Gambling Task. Schizophrenia research 72, 215-224 (2005).
98. Garon, N., Moore, C. & Waschbusch, D. A. Decision making in children with ADHD only, ADHD-anxious/depressed, and control children using a child version of the Iowa Gambling Task. Journal of Attention Disorders 9, 607-619 (2006).
99. De Martino, B., Harrison, N. A., Knafo, S., Bird, G. & Dolan, R. J. Explaining enhanced logical consistency during decision making in autism. Journal of Neuroscience 28, 10746-10750 (2008).
100. Noel, X., Brevers, D. & Bechara, A. A neurocognitive approach to understanding the neurobiology of addiction. Current opinion in neurobiology 23, 632-638 (2013).
101. Fineberg, N. A. et al. New developments in human neurocognition: clinical, genetic, and brain imaging correlates of impulsivity and compulsivity. CNS spectrums 19, 69-89 (2014).
102. Huys, Q. J., Daw, N. D. & Dayan, P. Depression: a decision-theoretic analysis. Annu Rev Neurosci 38, 1-23 (2015).
103. Frank, M. J., Seeberger, L. C. & O'Reilly R, C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306, 1940-1943 (2004).
104. Collins, A. G. E. & Frank, M. J. Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychological review 121, 337 (2014).
105. Cools, R. Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. Neurosci Biobehav Rev 30, 1-23 (2006).
106. Cools, R., Altamirano, L. & D'Esposito, M. Reversal learning in Parkinson's disease depends on medication status and outcome valence. Neuropsychologia 44, 1663-1673 (2006).
107. Cools, R. et al. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. The Journal of neuroscience 29, 1538-1543 (2009).
108. Abler, B., Greenhouse, I., Ongur, D., Walter, H. & Heckers, S. Abnormal reward system activation in mania. Neuropsychopharmacology 33, 2217-2227 (2008).
109. Bishop, S. J. & Gagne, C. Anxiety, Depression, and Decision Making: A Computational Perspective. Annual review of neuroscience (2018).

CHAPTER 2

A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states

Jeroen P.H. Verharen
Johannes W. de Jong
Theresia J.M. Roelofs
Christiaan F.M. Huffels
Ruud van Zessen
Mieneke C.M. Luijendijk
Ralph Hamelink
Ingo Willuhn
Hanneke E.M. den Ouden
Geoffrey van der Plasse
Roger A.H. Adan*
Louk J.M.J. Vanderschuren*

* Equal contribution

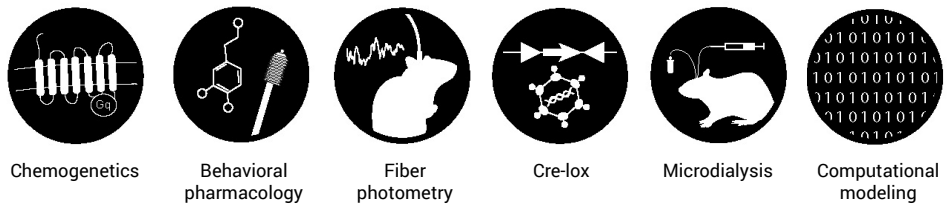
Published in Nature Communications 9: 731 (2018)

This study has been awarded the Rudolf Magnus Research Award 2018

Highlights

- An abundance of dopamine in the brain leads to over-optimistic and reckless choice behavior
- We show that this may be driven by impaired learning from negative feedback
- This learning impairment arises from impaired processing of negative reward prediction error signals in the nucleus accumbens

Techniques



CHAPTER 2

Hyperdopaminergic states in mental disorders are associated with disruptive deficits in decision-making. However, the precise contribution of topographically distinct mesencephalic dopamine pathways to decision-making processes remains elusive. Here we show, using a multidisciplinary approach, how hyperactivity of ascending projections from the ventral tegmental area (VTA) contributes to impaired flexible decision-making in rats. Activation of the VTA-nucleus accumbens pathway leads to insensitivity to loss and punishment due to impaired processing of negative reward prediction errors. In contrast, activation of the VTA-prefrontal cortex pathway promotes risky decision-making without affecting the ability to choose the economically most beneficial option. Together, these findings show how malfunction of ascending VTA projections affects value-based decision-making, providing a possible mechanism through which an abundance of dopamine may lead to aberrations in behavior, as is seen in substance abuse, mania, and after dopamine replacement therapy in Parkinson's disease.

Introduction

Impaired decision-making can have profound negative consequences, both in the short and in the long term. As such, it is observed in a variety of mental disorders, such as mania[1,2], substance addiction[3-6], and as a side effect of dopamine (DA) replacement therapy in Parkinson's disease[7,8]. Importantly, these disorders are associated with aberrations in DAergic neurotransmission[9,10], and DA has been implicated in decision-making processes[11-13]. However, ascending DAergic projections from the ventral mesencephalon are anatomically and functionally heterogeneous[14-16] and the contribution of these distinct DA pathways to decision-making processes remains elusive.

The mesocorticolimbic system, comprising DA cells within the ventral tegmental area (VTA) that mainly project to the nucleus accumbens (NAc; mesoaccumbens pathway) and medial prefrontal cortex (mPFC; mesocortical pathway), has an important role in value-based learning and decision-making[14-16]. When an experienced reward is better than expected, the firing of VTA DA neurons increases, thereby signaling a discrepancy between anticipated and experienced reward to downstream regions. Conversely, when a reward does not fulfill expectations, DA neuronal activity decreases. This pattern of DA cell activity is the basis of reward prediction error (RPE) theory[17-20], which describes an essential mechanism through which organisms learn to flexibly alter their behavior when the costs and benefits associated with different courses of action shift. Although the relevance of RPEs in value-based learning is widely acknowledged, little is known about how different VTA target regions process these DA-mediated error signals, and how this ultimately leads to adaptations in behavior.

Here, we used projection-specific chemogenetics combined with behavioral tasks, pharmacological interventions, computational modelling, *in vivo* microdialysis and *in vivo* neuronal population recordings to investigate how different ascending VTA projections contribute to value-based decision-making processes in the rat. Specifically, we investigated the mechanism underlying the aberrant decision-making style that is associated with increased DA neuron activity. We hypothesized that hyperactivation of VTA neurons interferes with reward prediction error processing, leading to impaired adaptation to reward value dynamics. We predicted an important contribution of the mesoaccumbens pathway in incorporating experienced reward, loss and punishment into future decisions, considering the importance of the NAc in reinforcement learning and motivated behaviors[21-23], and a

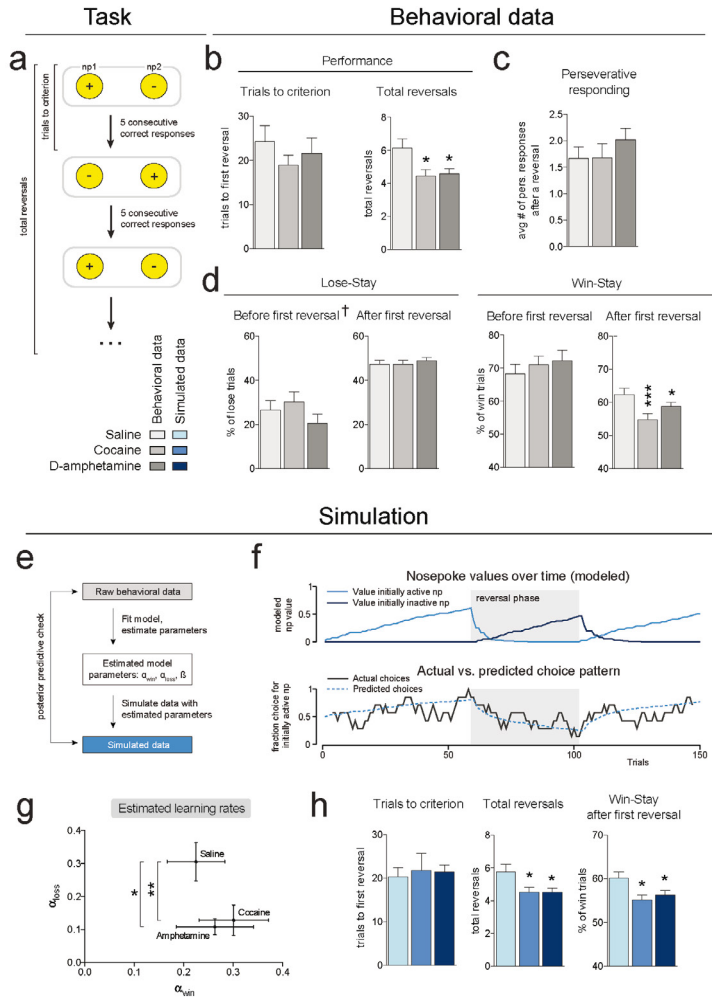


Figure 1 Treatment with cocaine or D-amphetamine impairs reversal learning. **(a)** Task design. **(b)** Systemic treatment with cocaine (10 mg/kg) or D-amphetamine (0.25 mg/kg) did not alter the number of trials required to reach the first reversal (1-way RM ANOVA, p = 0.55), but decreased the total number of reversals accomplished (ANOVA, p = 0.0037; post-hoc Sidak's test, p = 0.0102 cocaine vs. saline, p = 0.0197 D-amphetamine vs. saline). **(c)** Treatment with cocaine or D-amphetamine did not alter perseverative behavior (p = 0.46). **(d)** Lose-stay behavior was unaffected after cocaine and D-amphetamine treatment, both before (p = 0.21 †) and after (p = 0.77) first reversal. Cocaine and D-amphetamine decreased win-stay behavior after (ANOVA, p = 0.0007; post-hoc Sidak's test, p = 0.0009 for cocaine vs. saline, p = 0.0336, D-amphetamine vs. saline), but not before the first reversal (p = 0.67). Data in (b),(c),(d) and (g): repeated measures from n = 25 animals. † 6 animals had no losses before the first reversal, so the ANOVA was performed on data of n = 19 animals; graph shows n = 25. **(e)** We used a modified Rescorla-Wagner model to describe the behavior of the rats during reversal learning. **(f)** Simulated data from an example session. (upper panel) Simulated values of the nose pokes, given the rat's optimal model parameters and observed choices. (lower panel) Modeled choice probabilities, converted from the simulated nosepoke values using a softmax (unsmoothed), and the rat's actual choice pattern (smoothed over 7 trials). **(g)** Best-fit learning parameters. Treatment with cocaine and D-amphetamine significantly decreased α_{loss}, without affecting the other model coefficients. (Wilcoxon signed rank test, * p = 0.032, ** p = 0.0046, see Table S2) **(h)** Simulating data with the model parameters extracted in (g) replicated the drug-induced effects of the behavioral data shown in (b) and (d). (n = 25 simulated rats). Data are shown as mean ± s.e.m.

modulatory role for the mesocortical pathway in value-based choice behavior, given its involvement in executive functions, such as decision-making and behavioral flexibility[24,25]. Furthermore, we tested an explicit prediction based on a neurocomputational model of the DA system, in which impaired negative RPE processing is involved in learning deficits during DA replacement therapy[7,26].

Results

Dopaminomimetic drugs impair serial reversal learning

To test the role of DA in flexible value-based decision-making, rats were tested in a serial reversal learning task following systemic treatment with the DA neurotransmission enhancers cocaine and D-amphetamine. A reversal learning session (Fig. 1a) comprised 150 trials, and started with the illumination of two nose poke holes in an operant conditioning chamber. One of these was randomly assigned as active, and responding in this hole resulted in sucrose delivery under a fixed-ratio (FR) 1 schedule of reinforcement. When animals had made five consecutive correct responses, the contingencies reversed so that the previously inactive hole now became active, and vice versa.

Injection of either drug did not affect the number of trials needed to reach the criterion of a series of five consecutive correct responses (Fig. 1b, left panel). However, the number of reversals achieved in the entire session was significantly reduced in the drug-treated animals (Fig. 1b, right panel, and Fig. S1a). Thus, cocaine and D-amphetamine impaired task performance, but this effect did not appear until the moment of first reversal. We reasoned that this pre- and post-reversal segregation in drug effects on task performance is related to the structure of the task (Fig. 1a). That is, after every reversal, the value of the outcome of responding in the previously active hole declines, and conversely, the value associated with responding in the previously inactive hole increases. Accordingly, this task entails a combination of devaluation and revaluation mechanisms following reversals.

To understand the nature of the drug-induced deficit in reversal learning performance, we analyzed the animals' behavior in more detail. Perseverative responding, i.e. the average number of responses in the previously active hole directly after a reversal, was not altered after cocaine or D-amphetamine treatment (Fig. 1c). Lose-stay behavior, i.e., the percentage of (unrewarded) trials in the inactive nose poke hole followed by a response in the (still) inactive hole, was also not affected (Fig. 1d, left panel). However, win-stay behavior, i.e., the percentage of responses in the active nose poke hole after which the animal responded in that same active hole, was significantly decreased after treatment with cocaine or D-amphetamine (Fig. 1d, right panel). This drug-induced reduction in win-stay behavior indicates that even though the animals received a reward after responding in the active nose poke hole, they next sampled the inactive hole more often than after saline treatment. Importantly, win-stay behavior was only reduced after reversal, indicating that behavioral impairments were not the result of a general decline in task performance or sensitivity to reward.

Overall, the effects in the reversal learning task indicate that increased DA signaling after cocaine or D-amphetamine treatment did not impair the animals' ability to find the active nose poke hole at task initiation, hence to assign positive value to an action. Yet, when the values of (the outcome of) two similar actions (that is, responding in a nose poke hole) changed relative to each other, drug-treated animals were impaired in adjusting behavior, perhaps as a result of a valuation deficit. This suggests that treatment with these drugs disrupted the process of integrating recent wins or losses (i.e., a revaluation or a devaluation impairment, respectively) in decisions.

To gain insight into the mechanisms underlying impaired reversal learning, we modelled the behavior of each subject by fitting the data to a computational reinforcement learning model (Fig. 1e,f and Table S1). We used an extended version of the Rescorla-Wagner model[27,28], using two different learning rates, α_{win} and α_{loss}, describing the animal's ability to learn from wins and losses, respectively[29]. Such a model-based approach investigates task performance based on an extended history of trial outcomes, and not merely the most recent outcome, such as win- and lose-stay measures do, providing a more in-depth analysis of the learning capacity of the animals.

When comparing the Rescorla-Wagner model coefficients of the animals after saline with those after cocaine and D-amphetamine treatment, we observed a strong decrease in parameter α_{loss} without affecting α_{win} or choice stochasticity factor β (Fig. 1g,h, Fig. S1b,c and Table S2). This indicates that cocaine and D-amphetamine interfere with learning from negative, but not positive, RPEs.

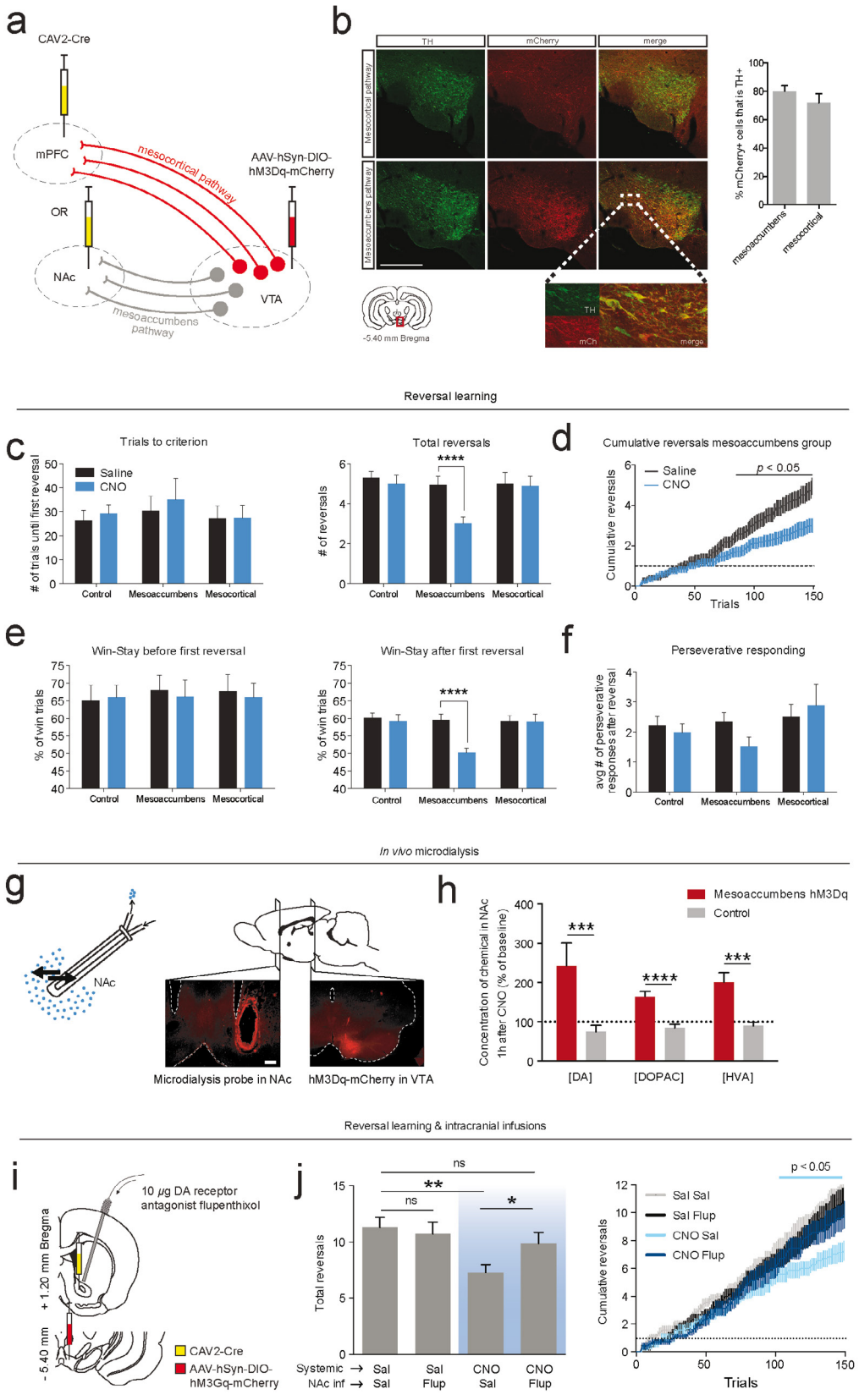
Chemogenetic activation of mesoaccumbens pathway impairs reversal learning

In view of the role of DA in RPE signaling, we hypothesized that cocaine and D-amphetamine interfered with learning from losses by overactivation of ascending midbrain DA projections, thereby disrupting negative RPEs. This same mechanism has been hypothesized to be involved in the DA dysregulation syndrome in medicated Parkinson's disease patients[7,30]. Such an overactivation may lead to an inability to devalue stimuli and/or their associated outcomes, resulting in choice behavior that is not optimally value-based. Specifically, we were interested in the contribution of projections from the VTA to the NAc and the mPFC to impairments in reversal learning.

In order to activate neuronal subpopulations of the VTA in a projection-specific manner, we combined a canine adeno-associated virus retrogradely delivering Cre-recombinase (CAV2-Cre) and a Cre-dependent viral vector encoding hM3Dq(Gq)-DREADD fused to mCherry-fluorescent protein[31] (Fig. 2a and Fig. S2). This two-viral approach resulted in high levels of DA specificity (80% of the transfected neurons in the mesoaccumbens group and 72% of the transfected neurons in the mesocortical group were positive for tyrosine hydroxylase, Fig. 2b). To investigate whether the effects of cocaine and D-amphetamine on reversal learning were driven by activation of the mesoaccumbens or mesocortical pathway, animals were injected with clozapine-N-oxide (CNO) immediately before testing in the reversal learning task.

Chemogenetic activation of the mesoaccumbens pathway resulted in the same pattern of impairments in reversal learning as cocaine and D-amphetamine treatment, i.e., a reduction in the numbers of reversals achieved, without affecting trials to first reversal criterion (Fig. 2c). This pattern was confirmed by plotting the cumulative reversals as a function of completed trials (Fig. 2d and Fig. S3a). Similar to cocaine and D-amphetamine, the performance impairment during mesoaccumbens activation was associated with a post-reversal (but not pre-reversal) decrease in win-stay behavior (Fig. 2e), whereas perseverative responding and lose-stay behavior were not altered (Fig. 2f and Fig. S3b). Remarkably, during mesoaccumbens activation, both win- and lose-stay behavior were around 50% post-reversal, indicative of random choice behavior. Indeed, the Rescorla-Wagner model fitted with a significantly lower likelihood after mesoaccumbens activation (Fig. S3c), indicating that the animals' performance declined such that the model was less able to describe the data compared to baseline conditions. In contrast to mesoaccumbens activation, mesocortical activation or CNO injection in a sham-operated control group had no effect on reversal learning.

The finding that hyperactivity in the mesoaccumbens pathway evoked similar effects on reversal learning as cocaine and D-amphetamine did, suggests that these drugs exert their influence on flexible value-based decision-making through DA neurotransmission within the NAc. To directly test this, we performed *in vivo* microdialysis in the NAc of animals that expressed Gq-DREADD in the mesoaccumbens pathway (Fig. 2g). Administration of CNO increased baseline levels of DA in the NAc, as well as its metabolites 3,4-dihydroxyphenylacetic acid (DOPAC) and homovanillic acid (HVA) (Fig. 2h and Fig. S4). Next, we infused the DA receptor antagonist α-flupenthixol into the NAc of DREADD-treated animals prior to



← **Figure 2** Chemogenetic activation of the mesoaccumbens, but not mesocortical pathway mimicked the effects of cocaine and D-amphetamine on reversal learning. **(a)** Experimental design. Animals received an infusion of CAV2-Cre into either the mPFC or NAc. A Cre-dependent Gq-DREADD virus was injected bilaterally into the VTA. **(b)** (left panel) Representative histology images showing coronal sections stained for tyrosine hydroxylase (left), DREADD-mCherry (middle) and an overlay (right). Image bottom left corner from Paxinos and Watson (2007). Scalebar, 500 µm. (right panel) Co-staining of mCherry with tyrosine hydroxylase, showing the percentage of DREADD-transfected neurons that is dopaminergic (mean ± s.d.). Data from n = 9 (mesoaccumbens), n = 8 (mesocortical) animals. **(c)** (left panel) Activation of either pathway did not affect the number of trials needed to reach the first reversal (i.e., 5 consecutive correct responses; two-way repeated measures ANOVA; main effect of CNO, p = 0.54; group × CNO interaction, p = 0.90). (right panel) Performance on the task over the entire session was significantly impaired after mesoaccumbens activation (two-way repeated measures ANOVA; main effect of CNO, p = 0.0025; group × CNO interaction, p = 0.0067; post-hoc Sidak's multiple comparisons test, p = 0.89 for control group, p < 0.0001 for mesoaccumbens group, p = 0.99 for mesocortical group) **(d)** Plot of the cumulative reversals over time shows that the performance deficit after mesoaccumbens activation does not appear until after the first reversal (Sidak's multiple comparisons test corrected for 150 comparisons, p < 0.05 after trial 85). Dashed line indicates first reversal. **(e)** A significant decrease in win-stay behavior after (two-way repeated measures ANOVA; main effect of CNO, p = 0.0040; group × CNO interaction, p = 0.0026; post-hoc Sidak's multiple comparisons test, p = 0.9647 for control group, p < 0.0001 for mesoaccumbens group, p = 0.9997 for mesocortical group), but not before first reversal (two-way repeated measures ANOVA; main effect of CNO, p = 0.78; group × CNO interaction, p = 0.91) was observed during mesoaccumbens activation. **(f)** Perseverative behavior was not affected (two-way repeated measures ANOVA; main effect of CNO, p = 0.89; group × CNO interaction, p = 0.71). All data: n = 17 control, n = 17 mesoaccumbens, n = 16 mesocortical group. **(g)** Microdialysis was used to measure extracellular concentrations of DA and its metabolites in the NAc after chemogenetic mesoaccumbens stimulation. Scalebar, 500 µm. **(h)** NAc levels of DA and its metabolites were elevated one hour after an i.p. CNO injection in DREADD-infected animals compared to controls (post-hoc tests, DA, p = 0.0002; DOPAC, p < 0.0001; HVA, p = 0.0008; see also Fig. S4). **(i)** Prior to reversal learning, animals received systemic CNO (or saline) for DREADD stimulation and a microinjection with α-flupenthixol (or saline) into the nucleus accumbens. **(j)** α-flupenthixol itself had no effect on reversal learning, but prevented the CNO-induced impairment on reversal learning (ANOVA, p = 0.0024; post-hoc Holm-Sidak's test: **p = 0.0019, *p = 0.0397). Note that animals had a higher baseline of reversals in this experiment, because the animals were trained on the task (see Online methods). Abbreviations: Sal, saline; Flup, α-flupenthixol; ns, not significant.

chemogenetic activation of the mesoaccumbens pathway in a reversal learning test (Fig. 2i). This dose of α-flupenthixol had no effect on reversal learning after systemic saline injection, but it restored the effect of chemogenetic activation of the mesoaccumbens pathway to a level statistically indistinguishable from saline treatment (Fig. 2j). This finding supports the assumption that the effects of mesoaccumbens hyperactivity are mediated through NAc DA receptor stimulation.

Dopamine neuron activity during reversal learning

Considering the function of RPEs in value updating[20], we tested whether midbrain DA neurons tracked the presence of wins and losses in the form of RPEs during reversal learning. To this aim, we measured *in vivo* neuronal population activity from DA neurons in the VTA using fiber photometry[32] in TH::Cre rats (Fig. 3a and Supplementary Movie 1).

Around the time of responding, we observed a clear two-component RPE signal[20] (Fig. 3b,c and Fig. S5), i.e. a ramping of DA activity towards the moment of response, followed by an additional value component. That is, win trials were associated with a prolonged DA peak, whereas loss trials were characterized by a rapid decline in DA population activity after the response was made. No such signals were observed in animals injected with an activity-

independent control fluorophore (Fig. S5).

Since mesoaccumbens hyperactivity only affected task performance after reversal, we compared DA activity pre- and post-reversal (Fig. 3c, right panels). In loss trials, we observed significantly stronger negative RPEs after the first reversal compared to before reversal. In contrast, DA peaks during the win trials were similar before and after the first reversal. This supports our notion that the impairment in reversal learning during mesoaccumbens hyperactivity was due to selective interference with learning from negative RPE-guided feedback.

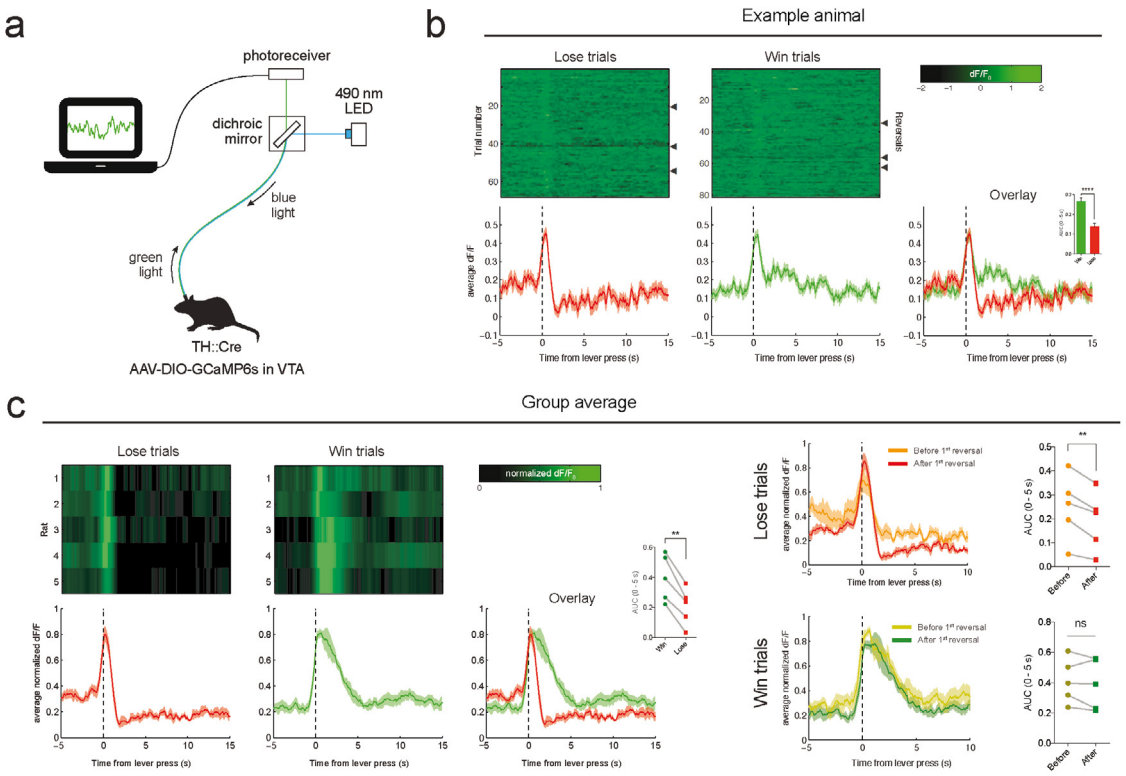


Figure 3 In vivo fiber photometry in VTA DA neurons during reversal learning.
(a) Experimental setup.
(b) Reversal learning session of an example animal. Triangles depict a reversal. Data is time-locked to a lever press by the rat and (in win trials) immediate reward delivery. Inset shows area under the curve in the first 5 seconds following lever press (unpaired t-test, $p < 0.0001$).
(c) Group average. (left panels) VTA DA neurons responded differentially to wins and losses (AUC (inset), paired t-test, $p = 0.0015$). (right panels) Lose trials evoked a stronger negative reward prediction error signal after the first reversal compared to before reversal. (AUC (inset), paired t-test, $p = 0.0062$ for lose trials, $p = 0.3658$ for win trials)

Mesoaccumbens hyperactivation interferes with adapting to devaluations

To examine whether the effects of mesoaccumbens hyperactivity on learning from negative feedback generalizes to conditions beyond reversal learning, we trained rats on a probabilistic discounting task (modified from refs. 33 and 34). In this task, rats could choose between responding on a 'safe' lever, which always produces one sucrose pellet, or on another, 'risky' lever, which produces a larger reward (i.e., three sucrose pellets) with a given probability. Within a session, the chance of receiving the large reward after a response on the risky lever decreases across four trial blocks — in the first block, animals always received the large reward when pressing the risky lever, whereas the odds of winning were reduced to 1 in 12 in the fourth block (Fig. 4a and Fig. S6a). An important difference with reversal learning is that in this task, a response shift is not the best option after a loss *per se* — lose-stay behavior at the risky lever may yield the same amount of sucrose as a shift to the safe lever, depending on the odds in the trials block. Therefore, an increase in lose-stay or decrease in win-stay behavior does not necessarily reflect poor choice behavior.

After training, the animals showed stable discounting performance, preferring the risky lever in the first block, and shifting their choice towards the safe lever when the yield of the risky lever diminished (Fig. 4b, left panel). Mesoaccumbens activation (Fig. 4b, middle panel) decreased the choice of the risky lever in the first block and increased choice for the risky lever in the last block, resulting in a significantly reduced slope of the discounting curve (Fig. 4b, middle panel, inset), and a lower percentage of optimal choices (Fig. 4c). Importantly, the inability to discount the value of the risky lever in the latter blocks of the task is indicative of an inability to adapt to a declining outcome of responding on the risky lever (Fig. S6b). The reduced choice for the risky lever in the first block may also be due to a devaluation deficit, as the receipt of only one sucrose pellet after responding on the safe lever (compared to the three pellet yield of responding on the risky lever) may be perceived as a 'loss', since the relative value of responding on the safe lever is lower in this block[35]. In contrast, mesocortical activation only increased risk-seeking in the second block, in which the yield of the safe (1 pellet) and risky (1 in 3 chance of 3 pellets) levers were equal (Fig. 4b, right panel), so that the amount of optimal choices remained unaffected (Fig. 4c). Further analysis of task strategy showed that lose-stay behavior at the risky lever was increased during activation of the mesoaccumbens and mesocortical pathways, whereas win-stay and safe-stay behavior were unaffected (Fig. 4d and Fig. S6c). Thus, activation of both ascending VTA projections made animals less prone to alter choice behavior after losses, which significantly impaired task performance during mesoaccumbens activation. The increase in lose-stay behavior during mesocortical activation is the result of the preference for the risky lever in the second trial block, but this did not result in poor choice behavior (Fig. 4c).

To test whether the effects in this task were specific to devaluation mechanisms, we trained the animals expressing DREADD in mesoaccumbens neurons on the same task with increasing, instead of decreasing odds of reward at the risky lever (Fig. 4e). In this condition, mesoaccumbens activation did not significantly change risky choice in any of the blocks (Fig. 4f), although a modest but significant decrease was observed in performance (i.e. a lower fraction of optimal choices; Fig. 4g) which was caused by a higher preference for the risky lever in the first few trials (Fig. S6d). This could be the result of a reduced ability of the animals to devalue the outcome of responding on the risky lever in the initial trials of the first block. However, since this version of the task primarily relies on revaluation, rather than devaluation mechanisms, especially in later blocks (Fig. S6b), a mesoaccumbens stimulation-induced devaluation deficit caused no further changes in behavior. Indeed, win-stay and lose-stay behavior were unaffected by mesoaccumbens activation (Fig. 4g).

In sum, the effects of chemogenetic activation on the probabilistic discounting task support our hypothesis that mesoaccumbens activation results in an inability of animals to adapt behavior to lower-than-expected outcomes, which under physiological circumstances

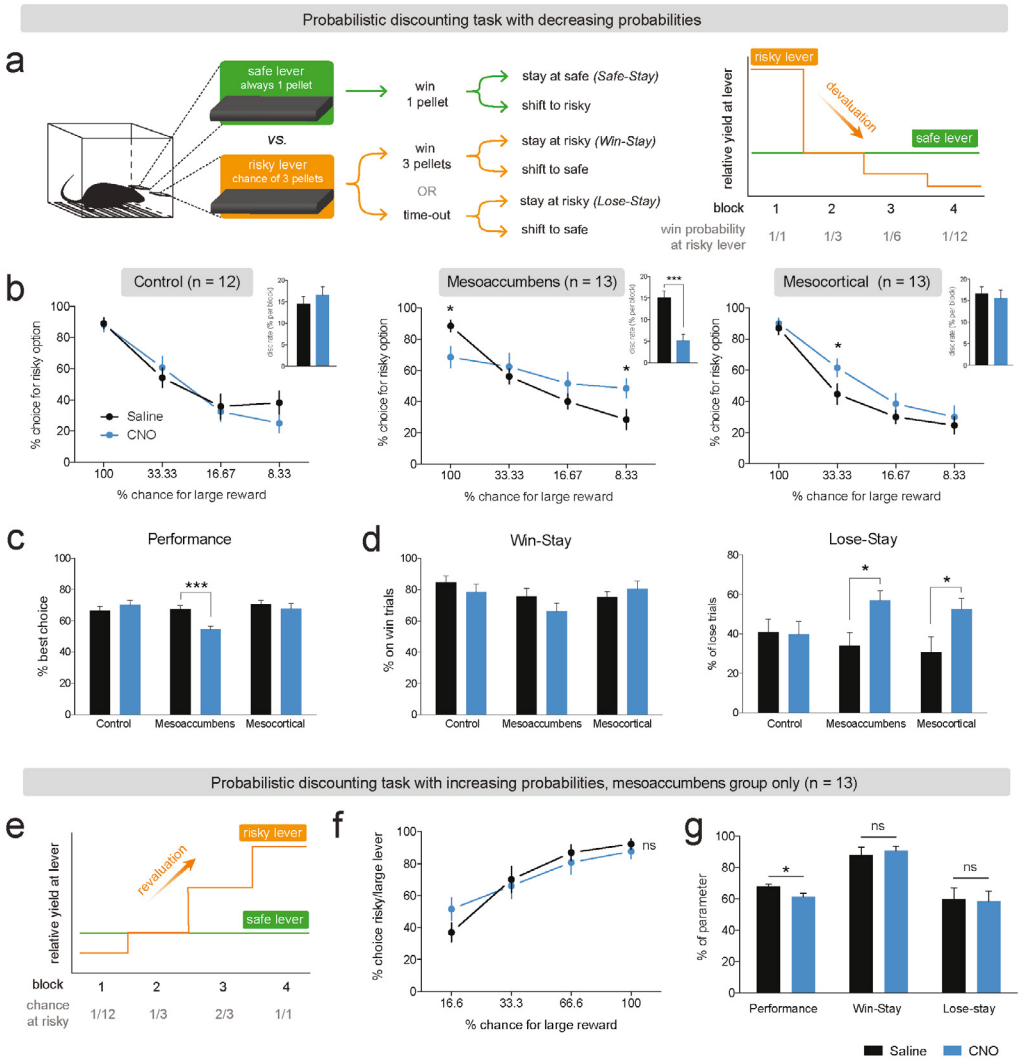


Figure 4 Chemogenetic activation of the mesoaccumens and the mesocortical pathway alters probabilistic discounting. **(a)** Task design. **(b)** Discounting curves for individual groups. (left panel) Sham control group (saline vs CNO; Sidak's test, $p > 0.1$ for all blocks). (middle panel) During mesoaccumbal hyperactivity, animals have a smaller preference for the risky lever in the first block (Sidak's test, $p = 0.0468$), a larger preference for the risky lever in the last block ($p = 0.0468$; block 2 and 3 both $p > 0.1$), and a significantly diminished discounting rate (inset, $p = 0.0002$). (right panel). Mesocortical activation increased choice for the risky lever in the second block (Sidak's test in block 2, $p = 0.0247$; block 1, 3 and 4, all $p > 0.1$). Asterisks in discounting curves indicate significant difference between saline and CNO treatment. Insets display the average steepness of the discounting curve (statistical comparison with Sidak's test). **(c)** Mesoaccumbens activation reduces the percentage optimal choices in the probabilistic discounting task (i.e., % best choice in blocks 1, 3 and 4; two-way repeated measures ANOVA; main effect of CNO, $p = 0.0331$; group × CNO interaction effect, $p = 0.0016$; post-hoc Sidak's test, $p = 0.5082$ for control group, $p = 0.0004$ for mesoaccumbens group, $p = 0.7533$ for mesocortical group). **(d)** Chemogenetic activation of the mesoaccumbens or mesocortical pathway had no effect on win-stay behavior (two-way repeated measures ANOVA; main effect of CNO, $p = 0.36$; group × CNO interaction

is mediated by negative RPE signals in DA cells. In contrast, mesoaccumbens hyperactivity did not markedly interfere with adaptations to higher-than-expected outcomes. Furthermore, mesocortical activation increased risky choice behavior, but only when this was without negative consequences for the net gain in the task.

Dopamine pathway activation does not change static reward value

Changes in static reward value may influence behavior in tasks investigating dynamic changes in reward value, such as the reversal learning task. For example, food rewards may be less or more appreciated due to changes in feelings of hunger, satiety or pleasure. Alternatively, operant responding may become habitual rather than goal-directed when manipulating the striatum, although this is thought to be mediated by its dorsal parts rather than the NAc[22,36].

To assess whether alterations in static reward value or in the associative structure of operant responding contributed to the behavioral changes evoked by DA pathway stimulation, rats were subjected to operant sessions in which they could lever press for sucrose under an FR-10 schedule of reinforcement. Activation of the mesoaccumbens and mesocortical pathways did not alter the total number of lever presses (Fig. 5a), suggesting that absolute reward value was unchanged. We also tested animals in operant sessions, whereby in half of the sessions the animals were pre-fed with the to-be obtained reward. This type of devaluation tests whether animals retain the capacity to adjust operant behavior to changes in (the representation of) reward value. Pre-feeding robustly diminished lever pressing for sucrose, both in a non-reinforced extinction session, as well as under an FR5 schedule of reinforcement. Importantly, this effect of chronic devaluation was not affected by mesoaccumbens or mesocortical activation (Fig. 5b), indicating that responding remained goal-directed[36].

Consistent with previous findings[37,38], activation of the mesoaccumbens pathway increased operant responding under a progressive ratio schedule of reinforcement[39] (Fig. 5c), which is often interpreted as reflecting an increased motivation to obtain food[37-39]. However, in light of the present findings, we interpret this finding to reflect that mesoaccumbens hyperactivity renders animals less able to devalue the relative outcome of pressing the active lever when the response requirement increases over the session, hence leading to increased response levels. Such an action devaluation likely involves negative RPE signals from DA neurons.

Mesoaccumbens hyperactivity evokes punishment insensitivity

To test whether the devaluation deficit as a result of mesoaccumbens hyperactivity also resulted in an inability to incorporate explicitly negative consequences into a decision, we subjected animals to a novel punishment task, in which reward taking was paired with an increasing chance of an inescapable footshock (Fig. 6a). As expected, the introduction of this 0.3 mA footshock punishment diminished responding for sucrose, an effect that persisted after injection of CNO in the mesocortical and sham control groups (Fig. 6b). In contrast, activation of the mesoaccumbens pathway completely abolished this punishment-

← effect, $p = 0.26$), but did increase lose-stay behavior (two-way repeated measures ANOVA; main effect of CNO, $p = 0.0026$; group × CNO interaction effect, $p = 0.0622$; post-hoc Sidak's test, $p = 0.9988$, $p = 0.0177$ and $p = 0.0203$ for control, mesoaccumbens and mesocortical groups, respectively). **(e)** Task design of the probabilistic discounting task with increasing probabilities. **(f)** Mesoaccumbens activation did not affect the discounting curve (Sidak's test in every block, $p > 0.1$). **(g)** Mesoaccumbens activation decreased performance on the task (paired t-test, $p = 0.0143$), but not win-stay (paired t-test, $p = 0.32$) or lose-stay behavior (paired t-test, $p = 0.85$). Data are shown as mean ± standard error of the mean.

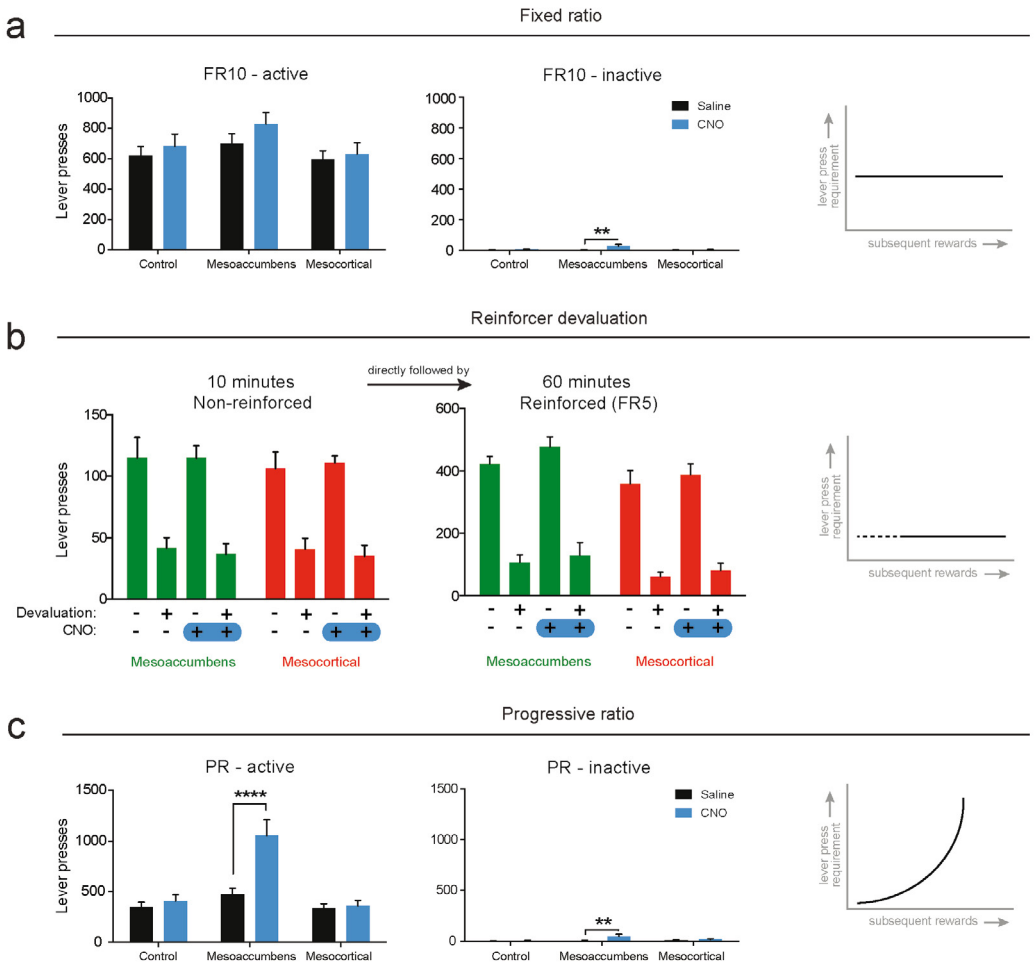


Figure 5 Mesocortical and mesoaccumbens activation does not alter the static reward value of sucrose. **(a)** DREADD activation of either pathway did not affect the number of active lever presses for sucrose under a fixed-ratio 10 schedule of reinforcement (two-way repeated measures ANOVA; main effect of CNO, $p = 0.0355$; group × CNO interaction, $p = 0.5001$; post-hoc Sidak's multiple comparisons test, CNO versus saline, all $p > 0.1$). n = 9 for control, n = 8 for mesoaccumbens group, n = 9 for mesocortical group. **(b)** Both during a 10-minute extinction session (left panel) and a reinforced lever pressing session (under an FR5 schedule of reinforcement, right panel), devaluation of the reinforcer by selective satiation for sucrose lead to a decrease in responding (2-way repeated measures ANOVA, main effect of prefeeding in all four groups, $p < 0.0001$), without any effects of CNO (non-reinforced mesoaccumbens, CNO effect $p = 0.7745$, prefeeding × CNO interaction: $p = 0.8448$; non-reinforced, mesocortical, CNO effect $p = 0.9516$, prefeeding × CNO interaction: $p = 0.5318$; reinforced mesoaccumbens, CNO effect $p = 0.1472$, prefeeding × CNO interaction: $p = 0.5287$; reinforced mesocortical, CNO effect $p = 0.4654$, prefeeding × CNO interaction: $p = 0.8877$). n = 12 for mesoaccumbens, n = 11 for mesocortical group. **(c)** Under a progressive ratio schedule of reinforcement, mesoaccumbens activation significantly increased the number of lever presses made (two-way repeated measures ANOVA; main effect of CNO, $p = 0.0006$; group × CNO interaction, $p = 0.0007$; post-hoc Sidak's multiple comparisons test, $p = 0.8998$ for controls; $p = 0.8998$ for control group; $p < 0.0001$ for mesoaccumbens group; $p = 0.9947$ for mesocortical group). n = 9 for control, n = 8 for mesoaccumbens group, n = 9 for mesocortical group. Data are shown as mean ± standard error of the mean.

induced reduction in responding, as the animals took as many rewards as under non-punishment conditions. This finding suggests that during mesoaccumbens hyperactivity, reward value is not properly discounted — in other words, animals are not able to take the increasingly negative consequences of an action into account. Consistent with a role for DA neurotransmission in processing these punishment signals, we observed, using *in vivo* calcium imaging, that footshock evoked a reduction in the activity of VTA DA neurons (Fig. 6c).

To control for effects of nociception in our punishment task, we subjected the animals to a tail withdrawal test, and found this not to be affected by mesoaccumbens activation (Fig. 6d). Moreover, anxiety, as tested in the elevated plus maze (Fig. S7a,b), was unaffected by mesoaccumbens stimulation. Consistent with literature, we found that mesoaccumbens stimulation increased locomotion (Fig. S8a), just like cocaine and D-amphetamine do[40,41]. We think, however, that the changes in value-based decision-making observed in the punishment task, as well as in the other tasks, cannot readily be attributed to increased locomotion. First, reaction times in the punishment task were longer after mesoaccumbens activation (Fig. S8b). Second, responding in the inactive hole in the punishment task was not changed (Fig. S8c). Third, the effects of mesoaccumbens activation in the reversal learning task were restricted to win-stay behavior after the first reversal. Last, mesoaccumbens activation did not affect the time for the animals to complete the reversal learning session (Fig. S3d).

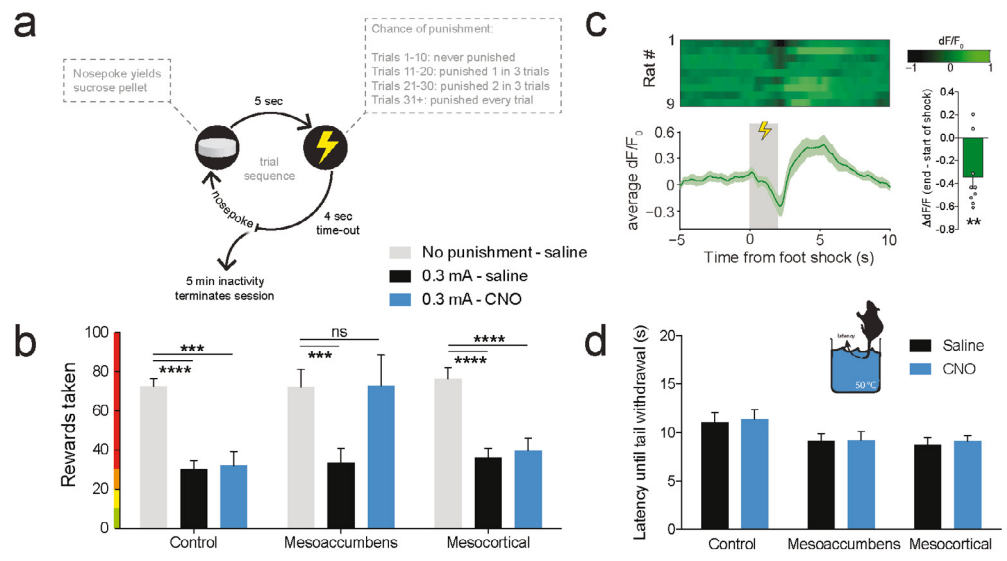


Figure 6 Mesoaccumbens, but not mesocortical activation attenuates the effect of punishment on responding for sucrose. **(a)** Task design. **(b)** After saline treatment, footshock punishment robustly diminished responding (Sidak's multiple comparisons test, '0.3 mA saline' versus 'no punishment saline', all $p < 0.001$). This effect was abolished by activation of the mesoaccumbens, but not the mesocortical, pathway (Sidak's test, '0.3 mA CNO' versus 'no punishment saline' in the mesoaccumbens group, $p = 0.9995$; in mesocortical group, $p = 0.0002$; in control group, $p < 0.0001$). n = 9 control, n = 9 mesoaccumbens group, n = 10 mesocortical group. **(c)** Footshock punishment evoked a decrease in DA neuron activity, measured using fiber photometry in TH::Cre rats (one-sample t-test, $p = 0.0074$, n = 9 rats). **(d)** No modulation of nociception by mesoaccumbens or mesocortical activation in the tail withdrawal test (2-way repeated measures ANOVA; main effect of CNO, $p = 0.75$; group × CNO interaction, $p = 0.99$). n = 8 control, n = 9 mesoaccumbens group, n = 9 mesocortical group. Data are shown as mean ± standard error of the mean. **** $p < 0.0001$, *** $p < 0.001$

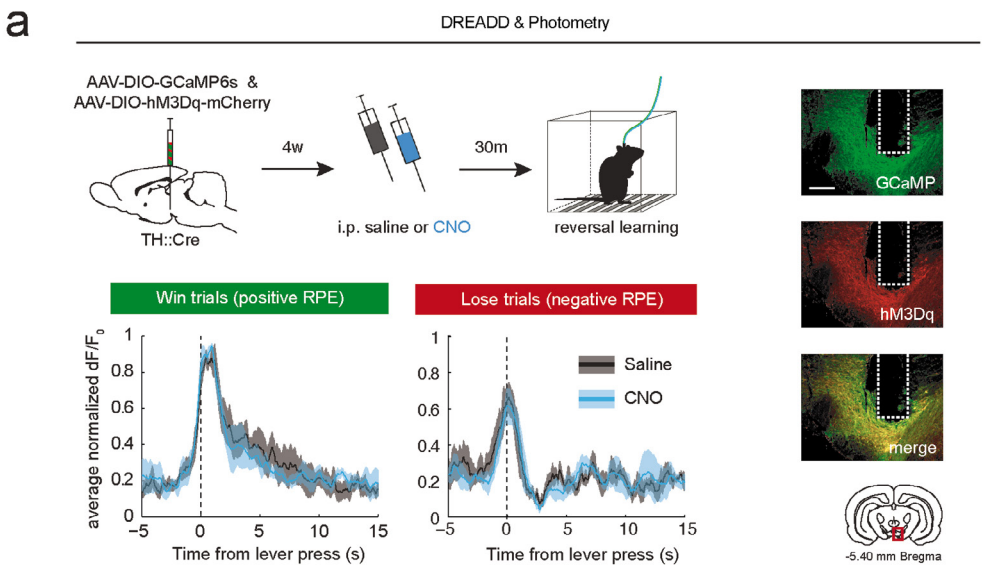


Figure 7 RPE processing after mesoaccumbens stimulation. **(a)** Animals were co-injected with GCaMP6s and Gq-DREADD and tested for reversal learning after injection of saline or CNO. VTA neurons responded in a comparable way during reversal learning after saline and CNO treatment (repeated measures in n = 4 animals; ANOVA, CNO x time interaction effect, win trials, p = 0.39; lose trials, p = 0.38). See figure S9a for individual animals. Scale bar, 1mm. Data are shown as mean (solid line) ± standard error of the mean (shading). **(b)** Proposed mechanisms: (I) Hyperactivity of NAc-projecting VTA DA neurons leads to impaired coding of negative RPE troughs, (II) Hyperactivity shifts baseline NAc DA levels, thereby preventing the exceedance of a negative RPE threshold in the NAc and impairing the ability to learn from negative feedback, or (III) A combination of both.

RPE processing during mesoaccumbens hyperactivity

There are three possible explanations for the impaired negative RPE processing during mesoaccumbens hyperactivity: (1) hyperactivity of VTA DA neurons abolishes the trough in neuronal activity caused by negative reward prediction, (2) elevated DA levels lead to a baseline shift in RPE signalling, after which a decrease in DA release during negative reward prediction does not reach the lower threshold necessary to provide a learning signal in downstream regions, or (3) a combination of both.

To address the first explanation, we unilaterally injected animals with a mixture of the calcium fluorophore GCaMP6s and Gq-DREADD and tested animals for reversal learning (Fig. 7a and Fig. S9). This allowed us to measure RPE signals from VTA neurons within one animal during baseline conditions and during hyperactivation of these same neurons. CNO administration did not impair the ability of VTA DA neurons to signal RPEs during reversal learning (i.e. deviations from baseline during reward prediction), inconsistent with the first possible explanation. By extension, this also excluded the third explanation. However, the second explanation is consistent with our findings that chemogenetic stimulation of the mesoaccumbens pathway increases the extracellular concentration of dopamine and its main metabolites in the NAc (Fig. 2h). Together, these data support a scenario in which the inability to adjust behavior after loss or punishment during hyperactivation of the mesoaccumbens pathway is not due to an inability of VTA neurons to decrease their firing rate during negative reward prediction, but rather by impaired processing of this learning signal within the NAc as a result of increased baseline DA levels (Fig. 7b). This observation fits well with our earlier finding that the infusion of a DA antagonist into the NAc can prevent the effects of DREADD activation on reversal learning (Fig. 2j), a manipulation that restores the degree of NAc DA receptor activation.

Discussion

Here, we show that hyperactivity of the mesoaccumbens pathway reduces the ability of animals to use loss and punishment signals to change behavior by interfering with negative RPE processing. Using *in vivo* neuronal population recordings, we show that the VTA signals reward presentation as well as reward omission during VTA neuron hyperactivity, meaning that the behavioral impairments are not caused by blunted DA neuron activity during negative reward prediction, but rather by impaired processing in the NAc as a result of elevated baseline levels of DA. Therefore, we propose a model (Fig. 7b) in which hyperactive VTA neurons signal positive and negative RPEs to the NAc, but because baseline DA tone is increased, the signaling threshold in the NAc that allows for the incorporation of negative RPEs into adaptive behavior cannot be reached during reward omission or punishment.

The majority of neurons transfected with the DREADD virus had a DAergic phenotype, chemogenetic mesoaccumbens activation replicated the effects of cocaine and D-amphetamine on reversal learning, and this effect of chemogenetic mesoaccumbens activation was prevented by intra-NAc infusion of the DA receptor antagonist α-flupenthixol. Together, this supports the notion that the behavioral changes observed in the present study are the result of chemogenetic stimulation of VTA DA cells. However, a role for non-DA neurons cannot be excluded with the currently used techniques. Importantly, alongside the dense DA innervation, the VTA sends GABAergic, glutamatergic, as well as mixed DA/GABA or DA/glutamate projections to the NAc and mPFC[16,42,43]. The role that these projections play in behavior is only beginning to be investigated, but on the basis of what is presently known, we consider it unlikely that the non-DAergic innervation of the NAc and mPFC is involved in the behavioral changes observed here. For example, optogenetic stimulation of VTA GABA neurons has been shown to suppress reward consumption, something we did not observe in our experiments[44]. In addition, by inhibiting NAc cholinergic interneurons, stimulation of VTA GABA projections to the NAc has been shown to enhance stimulus-outcome learning[45]. However, increased stimulus salience does not readily explain the deficits in

reversal learning, probabilistic discounting and punished responding for sucrose that we found in the present study. Last, stimulation of VTA-NAc glutamate neurons has been shown to produce aversive effects[46], which in our experiments most likely would have increased, rather than decreased the ability to use negative feedback to alter behavior. Therefore, we think it is justified to state that the deficits in reversal learning, probabilistic discounting and punished reward taking evoked by chemogenetic mesoaccumbens stimulation is the result of increased DA signaling in the NAc. Reversal learning impairments have previously been reported after systemic or intra-NAc treatment with a DA D_2 receptor agonist in rats and humans[47-49], whereas probabilistic discounting seems to be dependent on DA D_1 rather than D_2 receptor stimulation in the NAc[50]. Together, this suggests that the behavioral effects of mesoaccumbens hyperactivity observed here rely on stimulation of both DA receptor subtypes, depending on the task structure. Interestingly, the punishment insensitivity we observed after mesoaccumbens stimulation appears inconsistent with previous studies showing that treatment with amphetamine and the DA D2 receptor agonist bromocriptine make animals more sensitive to probabilistic punishment in a risky decision-making task, in which animals can choose between a small and safe reward, and a large reward with a chance of punishment[51,52]. In this latter task, however, presentation of the punishment coincides with the presentation of the large reward, and it is unknown how DA neurons respond to such an ambivalent combination of events. Importantly, risky choice behavior was found to correlate positively with DA D1 receptor expression in the NAc shell[52], suggesting that the influence of NAc DA on behavior in this task may not be unidirectional.

In contrast to the mesoaccumbens projection, hyperactivity of the mesocortical pathway did not markedly affect value-based decision-making. It did increase the preference for large, risky rewards over small, but safe rewards in the probabilistic discounting task. However, when one of the two options yielded more sucrose reward, animals remained capable of choosing the most beneficial option, perhaps as a result of the differential roles that prefrontal D1 and D2 receptors play in this task[53]. That these animals maintained the capacity to make proper value-based decisions was also apparent in the reversal learning and punishment tasks. Thus, the patterns of effects of mesocortical stimulation is qualitatively different from the mesoaccumbens-activated phenotype, even though there is modest overlap, such as the increased lose-stay behavior in the probabilistic discounting task. Therefore, we do not think that the mesocortical phenotype is an attenuated version of the mesoaccumbens one, although the lower density of the mesocortical projection (Fig. S2a) may explain the relative paucity of behavioural changes after chemogenetic mesocortical stimulation. Notably, the mesocortical pathway has been shown to be vital for certain forms of cost-benefit judgement, especially those involving uncertainty or sudden changes in task strategy[25]. As a result, manipulations of prefrontal DA affect tasks like probabilistic discounting or set shifting, but not reversal learning[25,54].

Our data emphasize the importance of balanced DA signaling in the NAc. It is reasonable to assume that brain DA concentrations are tuned to levels that are optimal to survival, and deviations from this optimum lead to the profound behavioral impairments seen in certain mental disorders. We think that our proposed model of mesoaccumbens overactivation can explain the decision-making deficits that are seen during states of increased DAergic tone, such as manic episodes, substance abuse, and DA replacement therapy in Parkinson's disease. When one cannot devalue stimuli, actions or outcomes based on negative feedback, their value representation remains artificially elevated. Hence, outcome expectancies of choices will be unrealistically high, leading to behavior that is overconfident and overoptimistic. These inflated outcome expectancies have been demonstrated in human manic patients[2], suggesting an inability to devalue goals towards realistic levels. That this disease state is associated with abolished negative RPE signaling in the NAc is substantiated by an fMRI study in patients experiencing acute mania[55], in which activity in the NAc of manic patients remained high when monetary reward was omitted, while healthy controls showed a significant reduction in NAc activity, as expected based on RPE theory.

Most drugs of abuse enhance DA transmission in the brain, either in a direct (e.g., DA reuptake inhibition) or indirect way (e.g., disinhibition of DA neurons)[56,57]. Direct dopaminomimetics, such as cocaine and D-amphetamine, are known to mimic the symptoms of mania, such as increased arousal, euphoria, and a reduced decision-making capacity[10]. Impaired learning from negative feedback may potentially contribute to the escalation of drug use, since users may be insensitive to the thought of forthcoming negative consequences during the 'high' of these drugs. Furthermore, DA replacement therapy, often prescribed to Parkinson's disease patients, has been associated with the development of problem gambling, hypersexuality and excessive shopping behavior, a phenomenon known as the DA dysregulation syndrome[58,59]. More than a decade ago, it has already been hypothesized that these clinical features could be the result of impaired RPE learning due to 'overdosing' midbrain DA levels[30,60]. Here, we provide direct evidence to support this notion.

There is a wealth of evidence to implicate increased DA levels in harmful decision-making behavior in mental disorders[1,2,3]. Thus far, however, it was unknown through which pathways and by which mechanisms these effects were mediated. Here, we used behavioral tasks in rats, combined with projection-specific chemogenetics to show that hyperactivation of the VTA leads to decision-making deficits by impairing negative feedback learning through overstimulation of NAc DA receptors. Altogether, we provide a mechanistic understanding of why decision-making goes awry during states of hyperdopaminergic tone, providing a possible explanation for the reckless behaviors seen during drug use, mania, and DA replacement therapy in Parkinson's disease.

Methods

Animals

A total of 128 adult male Crl:WU Wistar rats (Charles River, Germany) were used for the behavioral experiments, weighing ~250 gram at the start of the experiments. Rats were housed in pairs in a humidity- and temperature-controlled environment under a 12h:12h reversed day-night cycle (lights off at 7am). Rats in the photometry, microdialysis and intra-accumbens micro-infusion experiments were housed individually. Rats were food restricted (4g of normal chow per 100g body weight on test days, 5g per 100g body weight on remaining days) during the following experiments: reversal learning and probabilistic discounting. During the other behavioral tasks, animals had *ad libitum* access to standard chow (Special Diet Service, UK). Animals always had *ad libitum* access to water, except during behavioral tests. All experiments were approved by the Animal Ethics Committee of Utrecht University and conducted in agreement with Dutch laws (Wet op de Dierproeven, 1996) and European guidelines (Guideline 86/609/EEC).

Surgeries

Anaesthesia was induced with an i.m. injection of a mixture of 0.315 mg/kg fentanyl and 10 mg/kg fluanisone (Hypnorm, Janssen Pharmaceutica, Beerse, Belgium). Animals were placed in a stereotaxic apparatus (David Kopf Instruments, Tujunga, USA) and a small incision was made along the midline of the skull. One µl of CAV2-Cre virus (2.3×10^{12} particles/ml) was bilaterally injected into the NAc (+1.20 mm AP, ±2.80 mm ML from Bregma and -7.50 mm DV from the skull, at an angle of 10°) or the mPFC (+2.70 mm AP, ±1.40 mm ML from Bregma and -4.90 mm DV from the skull, at an angle of 10°). The control group received a bilateral injection of 1 µl saline into the NAc. All animals received a bilateral injection of 1 µl AAV5-hSyn-DIO-hM3Gq-mCherry (1×10^{12} particles/ml) into the VTA (-5.40 mm AP, ±2.20 mm ML from Bregma and -8.90 mm DV from the skull, at an angle of 10°). The viruses were infused at a rate of 0.2 µl/min. After injection, the needle was maintained at its injection position for 10 min to allow the virus to diffuse into the tissue. After surgery, the animals were given

carprofen for pain relief (5 mg/kg per day for 3 days, s.c.) and saline (10 ml once, s.c.). Animals were allowed to recover for 7 days before behavioral training continued. Behavioral testing started at least six weeks after surgery to allow for proper viral transfection.

Accumbens micro-infusions
For intra-accumbens micro-infusions, 7 animals were bilaterally implanted with 26-gauge stainless steel guide cannulas (Plastics One, Raonoke, USA), 1 mm above the NAc (same coordinates as for CAV2-Cre injection, see above), after injection of the viral vectors necessary for mesoaccumbens Gq-DREADD expression. Cannulas were secured to the skull by screws and dental cement. Injectors protruded 1 mm beyond the termination point of the guide cannulas.
Animals were habituated with saline infusions (0.5 µl/side) from 3 days before the experiment, 15 minutes before reversal learning training sessions. On the two experimental days, animals received infusions with saline (0.5 µl/side) or cis-(Z)-α-flupenthixol dihydrochloride (Sigma-Aldrich, Zwijndrecht, The Netherlands) dissolved in saline (10 µg dissolved in 0.5 µl/side), together with an i.p. injection of saline or CNO, 15 minutes prior to reversal learning. The infusion rate was set to 1 µl/min, and the injectors were left in place for an additional 30 seconds after the infusion was complete to allow for the diffusion of saline/flupenthixol into the brain. Between the time of infusion and testing, animals were placed back into their home cage.

Behavioral procedures
Animals were trained 5-7 days per week. All behavioral experiments took place between 9am and 6pm. Behavioral tests a-e (see below) were conducted in operant conditioning chambers (30.5×24.2×21.0 cm; Med Associates Inc., USA), placed within sound-attenuated cubicles. Experiments a, c and d were conducted in boxes that were equipped with a sucrose receptacle flanked by two retractable levers and cue lights. The wall on the other side of the box contained a house light and tone cue generator. Experiments b and e were conducted in different boxes that contained two illuminated nose pokes, a house light and a tone cue generator on one side of the box, and a sucrose receptacle flanked by two cue lights on the other side of the box. Sucrose pellets used were 45mg each (SP; 5TUL, TestDiet, USA).

Chemogenetic experiments were conducted in five independent cohorts of animals:
Cohort 1: Responding for sucrose: fixed ratio (FR) 5 schedule of reinforcement with prefeeding devaluation (with levers) [Fig. 5b], progressive ratio (PR) schedule of reinforcement (with levers) [Fig. 5c], open field [Fig. S8a]
Cohort 2: Responding for sucrose: FR 10 schedule of reinforcement (with levers) [Fig. 5a], elevated plus maze [Fig. S7]
Cohort 3: Probabilistic discounting (with levers) [Fig. 4], reversal learning (with nose pokes) [Fig. 2], punishment task (with nose pokes) [Fig. 6]
Cohort 4: Probabilistic discounting (with levers) [Fig. 4], reversal learning (with nose pokes) [Fig. 2], elevated plus maze [Fig. S7], tail withdrawal test [Fig. 6d]
Cohort 5: Probabilistic discounting (with levers) [Fig. 4f,g], reversal learning (with nose pokes) [Fig. 2i,j

CNO (0.3 mg/kg dissolved in 0.3 mg/ml saline) or saline was injected i.p., 20-30 minutes before the start of every experiment. Unless otherwise indicated, animals were treated with CNO and saline counterbalanced between days. In between treatment days, a wash-out period of at least 48 hours was used, during which behavioral training was continued.

a. Fixed ratio and progressive ratio schedule of reinforcement
Operant sessions under a fixed ratio (FR) schedule of reinforcement lasted for 1 hour, during which the house light was illuminated to signal response-contingent reward availability. Animals were first trained under an FR1 schedule of reinforcement, during which pressing the active lever resulted in the delivery of one sucrose pellet, the illumination of the cue light above the active lever for 5 s and retraction of both levers. After a 10-s time-out period (during which the house light was turned off), the levers were reintroduced and the house light was turned on, signaling the start of a new trial. Pressing the inactive lever was without scheduled consequences. After acquisition of sucrose self-administration under an FR 1 schedule, the response requirement was increased to FR5 (see below, experiment c), or FR10.
Under the progressive ratio (PR) schedule of reinforcement, the response requirement on the active lever was progressively increased after each obtained reward (1, 2, 4, 6, 9, 12, 15, 20, 25, etc., see ref. 61). A PR session ended after the animal failed to obtain a reward within 30 min. The animals were trained under FR and PR schedules before surgery. After surgery, they were retrained until we observed stable responding for at least 3 consecutive days at group level.

b. Reversal learning
Animals were trained to nose poke for sucrose under an FR1 schedule, in which responding in either of the two illuminated nose pokes resulted in the delivery of one sucrose pellet. During the reversal learning test, the nose poke holes were illuminated and responding into one of two holes (the site of the active hole was counterbalanced between animals) always resulted in reward delivery, a 0.5s auditory tone, and switching off the nose poke lights. Responding into the inactive hole always resulted in an 8s time-out period during which the house light and nose poke lights were turned off. A new trial began 8s after the last response, which was signaled to the animal by illumination of the nose poke lights. When the animal made 5 correct consecutive responses in the active hole, the contingencies were reversed so that the previously inactive hole became the active one, and the previously active hole became the inactive one. The session ended when the animal completed 150 trials.
Animals had no prior experience with contingency switches before the reversal learning experiments. In between treatment days, animals were retrained on an FR1 schedule of reinforcement, in which responding of any of the two nosepoke holes resulted in reward delivery. Before the intra-accumbens micro-infusions reversal learning experiments (Fig. 2i,j), animals received 8 reversal learning training sessions, to gain experience with contingency changes. This was done to minimize the chance on a between-days effect on performance, i.e. a difference in performance between the first and last testing day not caused by the manipulation.
Win-stay behavior was calculated as the percentage of rewarded trials on the active nose poke hole followed by a response on that same nose poke hole in the subsequent trial. Lose-stay behavior was calculated as the percentage of non-rewarded trials on the inactive nose poke hole after which the animal responded in that same nose poke hole in the subsequent trial. Trials to criterion was defined as the total number of trials necessary to reach the first reversal (i.e., 5 consecutive responses at the active nosepoke hole). Perseverative responding was defined as the total number of consecutive responses at the inactive nosepoke hole directly a reversal. For example, if after a reversal the animal chooses inactive-inactive-active, the # of perseverative responses after that reversal is 2.

c. Prefeeding devaluation
One hour before operant testing, animals were individually housed in standard cages where they had *ad libitum* access to water and standard chow (non-devalued situations) or sucrose pellets (devalued situation). The devaluation test comprised 10 minutes of non-reinforced lever pressing, during which pressing on either of the two levers was without scheduled consequences. This test was immediately followed by a regular session under an FR5

schedule of reinforcement. The animals were tested 4 times (devalued/non-devalued, CNO/saline), according to a within-subjects counterbalanced design. Each test day was followed by at least 2 days of regular FR5 training.

d. Probabilistic discounting task

This task was modified from refs. 33 and 34. Animals were allowed to respond on a safe lever, which always yielded one sucrose pellet, and a risky lever, which yielded three sucrose pellets with a given probability. The task comprised four blocks, each consisting of 6 forced trials on the risky lever (in which only the risky lever was presented), followed by 10 free choice trials (in which both the safe and the risky lever were presented). The chance of receiving a large reward at the risky lever decreased across the four trial blocks: 100%, 33%, 16.67% and 8.33% in blocks 1, 2, 3 and 4, respectively. Choosing the safe lever resulted in reward delivery (one pellet), a 0.5s audio tone and illumination of the cue light above the safe lever for 17s. Hereafter, an intertrial interval of 3s started, in which house- and cue light were turned off. A rewarded response on the risky lever started the same sequence of cues, except that three sucrose pellets were delivered, with an interval of 200 ms. A non-rewarded response on the risky lever resulted in a 20s time-out in which all lights in the operant chamber were turned off. A new trial was signaled by illumination of the house light and reintroduction of the levers. A switch of blocks was signaled to the animal by switching the houselight, cuelights and tone on and off within 2s (1s ON, 1s OFF), three times in a row. This was immediately followed by the start of the forced trials sequence.

Before training on the probabilistic discounting task, animals were trained to respond on both levers, in which one lever (the future safe lever) always yielded one sucrose pellet, and the other lever (the future risky lever) always three sucrose pellets. There were 3 trial types, each with a 33.3% probability: one in which only the single-pellet lever presented, one in which only the three-pellet lever was presented, and one in which both levers were presented so the rats could choose between either lever. Hereafter, animals were trained on the probabilistic discounting task until stable task performance was observed (no significant effect of training day in a repeated-measured ANOVA over 3 days).

Win-stay behavior was calculated as the percentage of rewarded trials on the risky lever followed by a response on the risky lever in the subsequent trial. Lose-stay behavior was calculated as the percentage of non-rewarded trials on the risky lever after which the animal responded on the risky lever in the subsequent trial. Performance was calculated as the % optimal choices in block 1, 3 and 4, thus % choice for the risky lever in block 1, and % choice for safe lever in blocks 3 and 4.

The discounting rate was calculated as follows:

$$\text{discounting rate (\% per block)} = \frac{\frac{p_{\text{block 3}}+p_{\text{block 4}}}{2}-p_{\text{block 1}}}{3} \quad (1)$$

With p being the percentage choice for the risky lever in the subscripted block. p_{block2} was left out of the equation because there is no economically best choice in the second block.

e. Punishment task

Animals were placed into the operant chamber and the session started with illumination of the house light and two nosepoke lights. Responding into the active nose poke hole resulted in the immediate delivery of one sucrose pellet, a 0.3s tone cue, and illumination of the cue lights on the other side of the operant chamber, next to the sucrose receptacle. House light and nose poke lights were turned off. Five seconds after the termination of the tone cue, a second 0.3s tone cue was played, which co-terminated with the chance of a 0.3 s, 0.3 mA foot shock. The chance of a foot shock increased across four trial blocks: trials 1-10, no punishment; trials 11-20, 1 in 3 trials punished; trials 21-30, 2 in 3 trials punished; trials 31 and up were always punished. Cue lights were turned off after the tone-foot shock combination terminated, leaving the animals in the dark during the 5s-inter-trial interval. Responding into the inactive hole was registered, but was without scheduled consequences. The session ended when no response into the active hole had been made for 5 min. Before animals were tested on the punishment task, animals were trained to nosepoke for sucrose under an FR1 schedule of reinforcement (i.e. the same task, but without foot shock punishment). Between the two testing sessions, animals were retrained to respond on FR1 (without punishment) for 2 days.

f. Tail withdrawal test

This test was modified from ref. 62. The animals were gently fixated in a towel and 3-5 cm of the tip of their tail was put in a beaker with water of 50 ± 1 °C. The latency until tail withdrawal was analyzed from a recorded video in a frame-by-frame manner. Animals were tested twice after CNO treatment, and twice after saline treatment (saline and CNO counterbalanced between days, with 48 hours in between). The latencies of the two respective tests were averaged. When the animal did not withdraw its tail within 20s, the animal was placed back into its home cage (this happened once in one animal).

g. Elevated plus maze

The elevated plus maze was made out of grey plexiglas, and consisted of two open arms (50×10 cm) and two closed arms (50×10×40 cm), connected by a center platform (10×10 cm). The maze was elevated 60 cm above the floor. Behavior was scored using Ethovision 3.0 (Noldus, Wageningen, The Netherlands). The total times spent in the closed arms, open arms, and on the central platform were analyzed. All animals received CNO and were tested once for 5 min to preserve novelty.

h. Open field test

The open field was 100×100 cm and made out of dark plexiglas. During the 5-minute test, the open field was illuminated with white light, and a white noise sound source (85 dB) was used to prevent distraction from ambient noise. Locomotor activity was measured using video tracking software (Ethovision 3.0, Noldus, Wageningen, The Netherlands). All animals received CNO and were tested once.

Computational model

To model the behavior of the animals in the reversal learning task, we fit the data to an extended Q-learning model. In this model, animal behavior is captured in three parameters:

- α_{win}: learning from positive RPE (win trials)
- α_{loss}: learning from negative RPE (lose trials)
- β: the extent to which choice behavior is driven by value

This model was chosen because it has a direct relation to midbrain dopamine by including reward prediction error factors in the equations.

On each trial, the value of left (Q_{left}) or right (Q_{right}) nose poke was updated, depending of which of those was chosen, according to the equation:

$$Q_{s,t} = \begin{cases} Q_{s,t-1} + \alpha_{win} \cdot RPE_{t-1} & \text{for win trials} \\ Q_{s,t-1} + \alpha_{loss} \cdot RPE_{t-1} & \text{for lose trials} \end{cases} \quad (2)$$

with

$$RPE_{t-1} = \begin{cases} 1 - Q_{s,t-1} & \text{for win trials} \\ 0 - Q_{s,t-1} & \text{for lose trials} \end{cases} \quad (3)$$

in which $Q_{s,t}$ is the value of the outcome of responding into nose poke *s* on trial *t*. Note that nose poke outcome values ranged from 0 to 1.

Nose poke outcome value at session start, $Q_{\text{left,t=1}}$ and $Q_{\text{right,t=1}}$,were set at 0.

Nose poke outcome values were converted to action probabilities using a softmax:

$$p_{s,t} = \frac{e^{\beta \cdot Q_{s,t}}}{e^{\beta \cdot Q_{left,t}} + e^{\beta \cdot Q_{right,t}}} \quad (4)$$

in which $p_{s,t}$ is the chance of choosing nose poke *s* in trial *t*.

Best-fit model parameters were determined per animal, per session by minimizing the model's negative log likelihood using MATLAB's 'fmincon' function. Each session's maximum likelihood was compared to a random choice model, in which every option had a 0.5 probability of being chosen, thus having had a log likelihood of 150 trials*log(0.5). The fit of the Rescorla Wagner model was compared with this random choice model, both on an individual level (Fig. S1b, S3c), and on a group level (Table S1), using a likelihood ratio test with the p threshold set at a liberal p = 0.1. This type of comparison is used, since the Rescorla Wagner model nests the chance model (chance model is a special case in the Rescorla Wagner model in which β = 0). Although some sessions were not well explained by the Rescorla Wagner model (i.e. animals chose randomly or used an alternative strategy; red dots in Fig. S1b, S3c), we decided to include all sessions in our between-treatment comparison to avoid a bias. Including only those animals in which all sessions were significantly better explained by the Rescorla Wagner model than by chance, resulted in the same effect (i.e. a decrease in α_{loss}), but with higher statistical significance.

The best-fit parameters for each condition (saline, cocaine, D-amphetamine) were compared within-animals, using a Wilcoxon matched-pairs signed rank test.

In vivo fiber photometry

Setup

A blue LED light (M490F2, Thorlabs, Germany) was coupled to a 400 µm core fiber optic patch cable (M76L01, Thorlabs) and connected to a fiber mount (F240FC-A, Thorlabs). It was then passed through an excitation filter (FF02-472/30-25, Semrock), reflected by a dichroic mirror (FF495/605-Di01-25x36, Semrock), and focused onto a 400 µm core (Made from BFH48-400, Thorlabs, CF440, Thorlabs) patch cable towards the animal. For in vivo experiments, this patch cable was connected to a 400 µm implantable fiber (BFH48-400, Thorlabs) using a 2.5 mm ceramic ferrule (CF440, Thorlabs). Returning green light passed through the same patch cable onto the fiber mount. It then passed through the dichroic mirror and was deflected by a second dichroic mirror (Di02-R594-25x36, Semrock, USA) and through an emission filter (FF01-535/50-25, Semrock). The light was then focused onto a silicon based photoreceiver (#2151 Photoreceiver, Newport corporation, USA) using a plano-concex lens (#62-561, Edmund Optics, USA).

After photo-electron conversion, the electrical signal was pre-amplified on the photodiode (2×10^{10} V/A or 2×10^{11} V/A) and then passed on to a lock-in amplifier (SR810, Stanford Research Systems). The lock-in amplifier was set to an AC grounded single input. It was then lock-in amplified in the range of 233-400 Hz, a 12 dB/oct bandwidth roll off and a 30 or 100 ms time constant for the subsequent low-pass filtering. Sensitivity settings of the detection ranged from 1 mV to 500 mV, with normal dynamic reserve and no additional notch filters applied. The lock in amplifier was set to the max offset (+109.21), and the phase was set to the hardware auto-adjusted value (typically in the range of 11-22 degree). The reference lock-in signal was translated by the hardware into TTL and coupled at 5V to the LED controller (LEDD1B, Thorlabs) that controlled the blue LED. The lock in amplified signal was then run onto an digitizer (Digidata 1550a Digitizer, Molecular Devices) and captured at 100 Hz – 10 kHz, typically using a 50Hz low pass filter. Additional TTL signals from behavioral events were simultaneously processed by the digitizer.

To correct for bleaching, raw data points F_x were converted to dF/F by running-average normalization:

$$(dF/F)_x = \frac{F_x - F_0}{F_0} \quad (5)$$

Here, F_0 is the baseline, which is calculated as the average of the 50% middle values in the 30 seconds following every time point F_x.

Experiment

The same surgical protocol as described above was used. Nine male TH::Cre rats (weighing 300-350 gram during surgery) were used, and 1 µl of AAV5-FLEX-hSyn-GCaMP6s or AAV5-hSyn-eYFP (University of Pennsylvania Vector Core) was injected at a titer of 10^{12} particles/ml unilaterally into the right VTA. A 400-µm implantable fiber was lowered to 0.1 mm above the injection site and attached with dental cement. Animals were tested in the reversal learning task described above, with the difference that retractable levers were used rather than nosepokes. This was done to prevent the dopamine transients to be influenced by perseverative responses into the nose poke during the inter-trial interval. Here, the levers remained retracted during the entire inter-trial interval, so that no responses could be made until the start of the next trial. In addition, no cue lights were used and the houselight was turned on continuously to prevent light contamination by the environment. Moreover, the correct responses in a row needed to obtain a reversal was set to 8 rather than 5, to increase the number of trials before the first reversal. Peri-stimulus time histograms were time-locked to the lever press (i.e., the moment of choice). In addition, 4 animals were injected with a 1-µl mixture of AAV5-FLEX-hSyn-GCaMP6s and AAV5-hSyn-DIO-hM3Gq-mCherry (both 10^{12} particles/ml, unilaterally in the right VTA). The AAV carrying Gq-DREADD was injected unilaterally in order not to interfere with task performance. Animals were tested in a counterbalanced fashion, so that half of the animals was first tested with saline, and the other half with CNO. In all animals expressing GCaMP6s, we tested whether a modest (0.30 mA) 2-second foot shock punishment evoked a negative RPE signal in VTA DA neurons. This was repeated 12 times in one session (with an inter-shock-interval of 40s).

Microdialysis

For microdialysis experiments, 9 animals were unilaterally implanted with guide cannulas (AgnTho's, Lidingö, Sweden), 1.5 mm above the right NAc (same coordinates as for CAV2-Cre injection, see above), 4 of which also received an injection of the viral vectors necessary for unilateral mesoaccumbens Gq-DREADD expression. After 4-6 weeks, a microdialysis probe (PES membrane protruding 2 mm beyond the cannula, cut-off 15 kD; AgnTho's,

Lidingö, Sweden) was placed into the guide cannula and secured. The following day, the microdialysis experiment commenced, by dialysing Ringer's solution through the probe at a rate of 1 µl/min. Each sample contained 15 µl of perfusate (i.e. 15 minutes), which was collected in 5 µl anti-oxidant solution containing 0.02 N HCOOH and 0.1% cysteine HCl in milli-Q. Saline, followed by CNO (1 mg/kg) was injected i.p., during dialysis.
Samples were analyzed by high performance liquid chromatography (HPLC) on an Alexis 100 2D system (ANTEC Leyden, Zoeterwoude, The Netherlands), at a flow rate of 0.035 ml/min. The mobile phase consisted a solution of 2.4 mM octanesulphonic acid, 1 mM KCl, 100 mM phosphoric acid and 15% methanol in milliQ. Chromatograms were analyzed using Clarity software (DataApex, Prague, Czech Republic).

Immunohistochemistry
Animals were euthanized by an i.p. injection of sodium pentobarbital and perfused with phosphate-buffered saline (PBS) followed by 4% paraformaldehyde (PFA) in PBS. The brains were dissected and postfixed in 4% PFA in PBS for 24 hours and then stored in a 30% sucrose in PBS solution. Brain slices (40 µm) were incubated overnight in a primary antibody solution, containing PBS with 0.3% Triton-X, 3% goat serum, and primary antibodies (1:1,000) against dsRed (rabbit, Clontech 632496) and TH (mouse, Millipore MAB318). The next day, brain slices were transferred to a secondary antibody solution containing PBS with 0.1% Triton-X, 3% goat serum, and secondary goat antibodies (1:1,000) against mouse (488 nm, Abcam ab150113) and rabbit (568 nm, Abcam ab175471). After an incubation period of 2hr at room temperature, slices were washed with PBS and mounted to glass slides. Histological verification was performed by a researcher unaware of the outcome of the behavioral experiment.

Exclusion criteria
Only animals that showed bilateral expression of hM3Gq-mCherry in the VTA were included in analyses. To exclude non-learners, animals in the probabilistic discounting task that showed a discounting rate of less than 10% per block at the end of training were excluded from the analysis.
Outlier analyses were performed on all data using the ROUT method (Q threshold set at 1.0%). Two rats were identified as outliers and removed from their respective datasets: one rat from the mesocortical group in the elevated plus maze experiment (outlier in time spent in closed arm), and one rat from the in vivo fiber photometry experiment on the basis of foot shock data (outlier in DA response to foot shock).

Data availability
The datasets generated during the current study are available from the corresponding author on reasonable request.

Code availability
Custom-written MATLAB and MedPC scripts are available upon request.

Data analysis and statistics
Data analysis and computational modelling was performed with MATLAB version R2014a (The MathWorks Inc.), statistical analyses with GraphPad Prism version 6.0 (GraphPad Software Inc.).
Statistical comparisons were made using a t-test for a single comparison, and a (repeated measures) ANOVA was used for multiple comparisons, followed by a t-test with Šidák's multiple comparisons correction. Paired, non-normally distributed data was compared using a Wilcoxon matched-pairs signed rank test with a Bonferroni correction for multiple comparisons. Welch's correction was used once, in a case where variances in the t-test were unequal.
Bar graphs represent the mean ± standard error of the mean, unless stated otherwise. In all figures: ^{ns} not significant, $^{\#}$ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001.

References

1. Murphy, F. C. *et al.* Decision-making cognition in mania and depression. *Psychol. Med.* **31**, 679-693 (2001).
2. Johnson, S. L. Mania and dysregulation in goal pursuit: a review. *Clin. Psychol. Rev.* **25**, 241-262 (2005).
3. Rogers, R. D. *et al.* Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: evidence for monoaminergic mechanisms. *Neuropsychopharmacology* **20**, 322-339 (1999).
4. Grant, S., Contoreggi, C. & London, E. D. Drug abusers show impaired performance in a laboratory test of decision making. *Neuropsychologia* **38**, 1180-1187 (2000).
5. Noel, X., Brevers, D. & Bechara, A. A neurocognitive approach to understanding the neurobiology of addiction. *Curr. Opin. Neurobiol.* **23**, 632-638 (2013).
6. Fineberg, N. A. *et al.* New developments in human neurocognition: clinical, genetic, and brain imaging correlates of impulsivity and compulsivity. *CNS Spectr.* **19**, 69-89 (2014).
7. Frank, M. J., Seeberger, L. C. & O'Reilly R, C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306**, 1940-1943 (2004).
8. Cools, R. Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. *Neurosci. Biobehav. Rev.* **30**, 1-23 (2006).
9. Volkow, N. D. & Morales, M. The brain on drugs: from reward to addiction. *Cell* **162**, 712-725 (2015).
10. van Enkhuizen, J. *et al.* The catecholaminergic-cholinergic balance hypothesis of bipolar disorder revisited. *Eur. J. Pharmacol.* **753**, 114-126 (2015).
11. Zeeb, F. D., Robbins, T. W. & Winstanley, C. A. Serotonergic and dopaminergic modulation of gambling behavior as assessed using a novel rat gambling task. *Neuropsychopharmacology* **34**, 2329-2343 (2009).
12. Linnet, J. *et al.* Striatal dopamine release codes uncertainty in pathological gambling. *Psychiatry Res.* **204**, 55-60 (2012).
13. Zalocusky, K. A. *et al.* Nucleus accumbens D2R cells signal prior outcomes and control risky decision-making. *Nature* **531**, 642-646 (2016).
14. Fields, H. L., Hjelmstad, G. O., Margolis, E. B. & Nicola, S. M. Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement. *Annu. Rev. Neurosci.* **30**, 289-316 (2007).
15. Lammel, S., Lim, B. K. & Malenka, R. C. Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology* **76**, 351-359 (2014).
16. Morales, M. & Margolis, E. B. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nat. Rev. Neurosci.* **18**, 73-85 (2017).
17. Schultz, W., Dayan, P. & Montague, P. R. A Neural Substrate of Prediction and Reward. *Science* **275**, 1593-1601 (1997).
18. Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* **16**, 966-973 (2013).
19. Keiflin, R. & Janak, P. H. Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. *Neuron* **88**, 247-263 (2015).
20. Schultz, W. Dopamine reward prediction-error signalling: a two-component response. *Nat. Rev. Neurosci.* **17**, 183-195 (2016).
21. Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.*

26, 321-352 (2002).
22. Voorn, P., Vanderschuren, L. J.M.J., Groenewegen, H. J., Robbins, T. W. & Pennartz, C. M. A. Putting a spin on the dorsal-ventral divide of the striatum. *Trends Neurosci.* **27**, 468-474 (2004).
23. Floresco, S. B. The nucleus accumbens: an interface between cognition, emotion, and action. *Annu. Rev. Psychol.* **66**, 25-52 (2015).
24. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167-202 (2001).
25. Floresco, S. B. Prefrontal dopamine and behavioral flexibility: shifting from an "inverted-U" toward a family of functions. *Front. Neurosci.* **7**, 62 (2013).
26. Collins, A. G. E. & Frank, M. J. Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev.* **121**, 337 (2014).
27. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current Research and Theory* **2**, 64-99 (1972).
28. Sutton, R. S. & Barto, A. G. Reinforcement learning: An introduction. (MIT press, 1998).
29. den Ouden, H. E. M. *et al.* Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* **80**, 1090-1100 (2013).
30. Cools, R., Barker, R. A., Sahakian, B. J. & Robbins, T. W. Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cereb. Cortex* **11**, 1136-1143 (2001).
31. Boender, A. J. *et al.* Combined use of the canine adenovirus-2 and DREADD-technology to activate specific neural pathways in vivo. *PLoS One* **9** (2014).
32. Gunaydin, L. A. *et al.* Natural neural projection dynamics underlying social behavior. *Cell* **157**, 1535-1551 (2014).
33. St Onge, J. R., Stopper, C. M., Zahm, D. S. & Floresco, S. B. Separate prefrontal-subcortical circuits mediate different components of risk-based decision making. *The J. Neurosci.* **32**, 2886-2899 (2012).
34. Cardinal, R. N. & Howes, N. J. Effects of lesions of the nucleus accumbens core on choice between small certain rewards and large uncertain rewards in rats. *BMC Neurosci.* **6**, 37 (2005).
35. Tobler, P. N., Fiorillo, C. D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307** (2005).
36. Yin, H. H., Ostlund, S. B. & Balleine, B. W. Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *Eur. J. Neurosci.* **28**, 1437-1448 (2008).
37. Zhang, M., Balmadrid, C. & Kelley, A. E. Nucleus accumbens opioid, GABAergic, and dopaminergic modulation of palatable food motivation: Contrasting effects revealed by a progressive ratio study in the rat. *Behav. Neurosci.* **117**, 202-211 (2003).
38. Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**, 470-485 (2012).
39. Hodos, W. Progressive ratio as a measure of reward strength. *Science* **134**, 943-944 (1961).
40. Beninger, R. J. The role of dopamine in locomotor activity and learning. *Brain Res. Rev.* **6**, 173-196 (1983).
41. Vanderschuren, L. J. M. J., Schoffelmeer, A. N. M., Wardeh, G. & De Vries, T. J. Dissociable effects of the kappa-opioid receptor agonists bremazocine, U69593, and U50488H on locomotor activity and long-term behavioral sensitization induced by amphetamine and cocaine. *Psychopharmacology* **150**, 35-44 (2000).
42. Van Bockstaele, E. J. & Pickel, V. M. GABA-containing neurons in the ventral tegmental area project to the nucleus accumbens in rat brain. *Brain Res.* **682**, 215-221 (1995).
43. Margolis, E. B. *et al.* Kappa opioids selectively control dopaminergic neurons projecting to the prefrontal cortex. *Proc. Natl. Acad. Sci. USA* **103**, 2938-2942 (2006).
44. van Zessen, R., Phillips, J. L., Budygin, E. A. & Stuber, G. D. Activation of VTA GABA neurons disrupts reward consumption. *Neuron* **73**, 1184-1194 (2012).
45. Brown, M. T. *et al.* Ventral tegmental area GABA projections pause accumbal cholinergic interneurons to enhance associative learning. *Nature* **492**, 452-456 (2012).
46. Qi, J. *et al.* VTA glutamatergic inputs to nucleus accumbens drive aversion by acting on GABAergic interneurons. *Nat. Neurosci.* **19**, 725-733 (2016).
47. Mehta, M. A., Swainson, R., Ogilvie, A. D., Sahakian, J. & Robbins, T. W. Improved short-term spatial memory but impaired reversal learning following the dopamine D2 agonist bromocriptine in human volunteers. *Psychopharmacology* **159**, 10-20 (2001).
48. Boulougouris, V., Castane, A. & Robbins, T. W. Dopamine D2/D3 receptor agonist quinpirole impairs spatial reversal learning in rats: investigation of D3 receptor involvement in persistent behavior. *Psychopharmacology* **202**, 611-620 (2009).
49. Haluk, D. M. & Floresco, S. B. Ventral striatal dopamine modulation of different forms of behavioral flexibility. *Neuropsychopharmacology* **34**, 2041-2052 (2009).
50. Stopper, C. M., Khayambashi, S. & Floresco, S. B. Receptor-specific modulation of risk-based decision making by nucleus accumbens dopamine. *Neuropsychopharmacology* **38**, 715-728 (2013).
51. Mitchell, M. R., Vokes, C. M., Blankenship, A. L., Simon, N. W. & Setlow, B. Effects of acute administration of nicotine, amphetamine, diazepam, morphine, and ethanol on risky decision-making in rats. *Psychopharmacology* **218**, 703-712 (2011).
52. Simon, N. W. et al. Dopaminergic modulation of risky decision-making. *J. Neurosci.* **31**, 17460-17470 (2011)
53. St Onge, J. R., Abhari, H. & Floresco, S. B. Dissociable contributions by prefrontal D1 and D2 receptors to risk-based decision making. *J. Neurosci.* **31**, 8625-8633 (2011).
54. Crofts, H. S. *et al.* Differential effects of 6-OHDA lesions of the frontal cortex and caudate nucleus on the ability to acquire an attentional set. *Cereb. Cortex* **11**, 1015-1026 (2001).
55. Abler, B., Greenhouse, I., Ongur, D., Walter, H. & Heckers, S. Abnormal reward system activation in mania. *Neuropsychopharmacology* **33**, 2217-2227 (2008).
56. Di Chiara, G. & Imperato, A. Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. *Proc. Natl. Acad. Sci. USA* **85**, 5274-5278 (1988).
57. Lüscher, C. & Ungless, M. A. The mechanistic classification of addictive drugs. *PLoS Med.* **3**, e437 (2006).
58. Evans, A. H. & Lees, A. J. Dopamine dysregulation syndrome in Parkinson's disease. *Curr. Opin. Neurol.* **17**, 393-398 (2004).
59. Berk, M. *et al.* Dopamine dysregulation syndrome: implications for a dopamine hypothesis of bipolar disorder. *Acta Psychiatr. Scand.* **116**, 41-49 (2007).
60. Frank, M. J. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J. Cogn. Neurosci.* **17**, 51-72 (2005).
61. Richardson, N. R. & Roberts, D. C. S. Progressive ratio schedules in drug self-administration studies in rats: a method to evaluate reinforcing efficacy. *J. of Neurosci. Methods* **66** (1996).

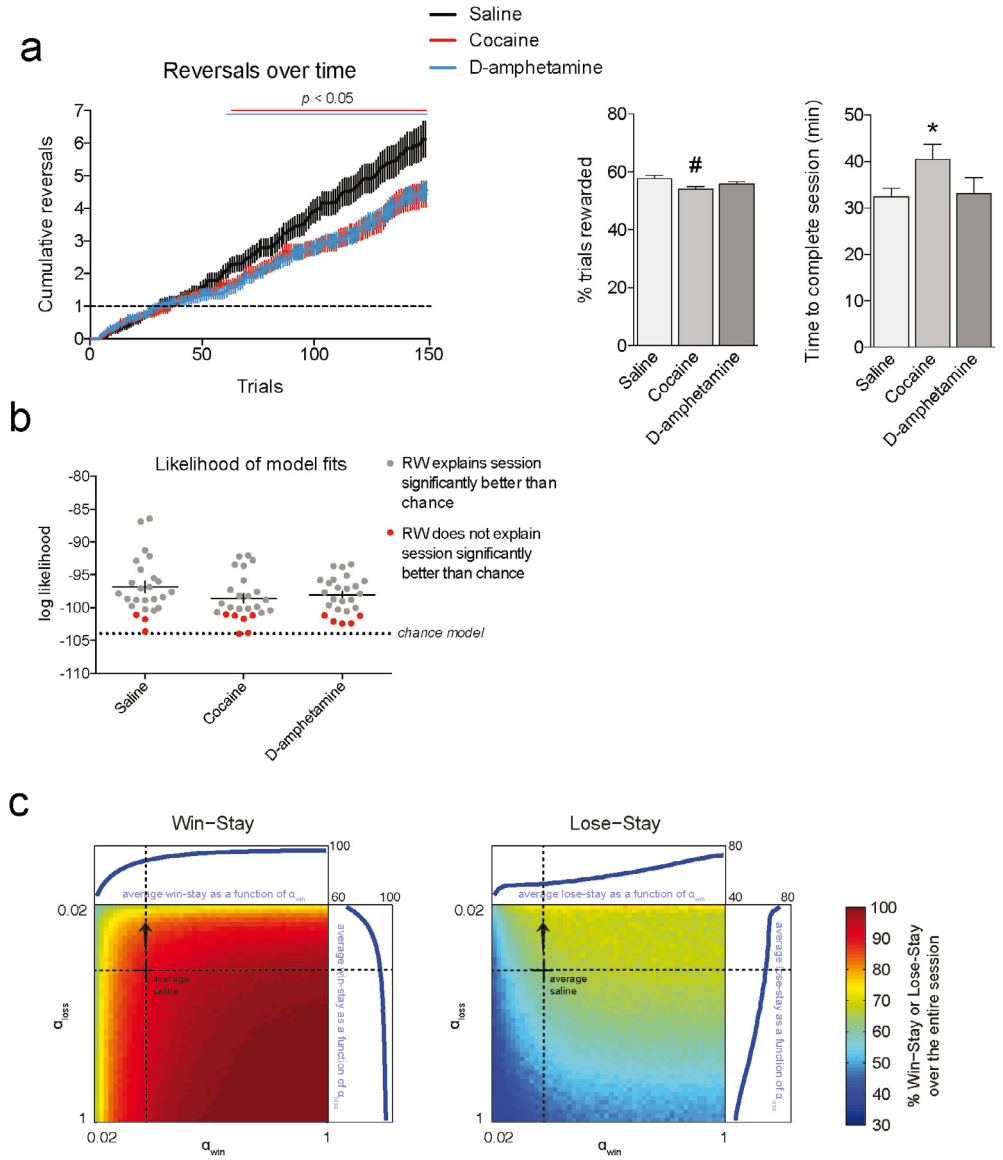
62. Nieh, E. H. *et al.* Decoding Neural Circuits that Control Compulsive Sucrose Seeking. *Cell* **160**, 528-541 (2015).

ACKNOWLEDGEMENTS
Clozapine-N-oxide was a generous gift from the NIMH Chemical Synthesis and Drug Supply Program. We thank Roshan Cools for giving feedback on the manuscript, and the entire Adan and Vanderschuren labs for helpful discussions and feedback. This work was supported by the European Union Seventh Framework Programme under grant agreement number 607310 (*Nudge-IT*), and the Netherlands Organisation for Scientific Research (NWO) under project numbers 912.14.093 (*Shining light on loss of control*) and 863.13.018 (*NWO/ALW Veni grant*).

AUTHOR CONTRIBUTIONS
J.P.H.V., J.W.D.J., G.v.d.P., R.A.H.A. and L.J.M.J.V. designed the experiments. J.P.H.V., J.W.D.J., T.J.M.R., C.F.M.H., R.v.Z., M.C.M.L., G.v.d.P. and R.H. performed the experiments. J.P.H.V. analyzed the behavioral and calcium imaging data. J.P.H.V. performed and H.E.M.d.O. supervised the computational analysis. I.W. and R.H. analyzed the microdialysis experiments. J.P.H.V., H.E.M.d.O., R.A.H.A. and L.J.M.J.V. wrote the paper with input from the other authors.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

SUPPLEMENTARY FIGURE 1

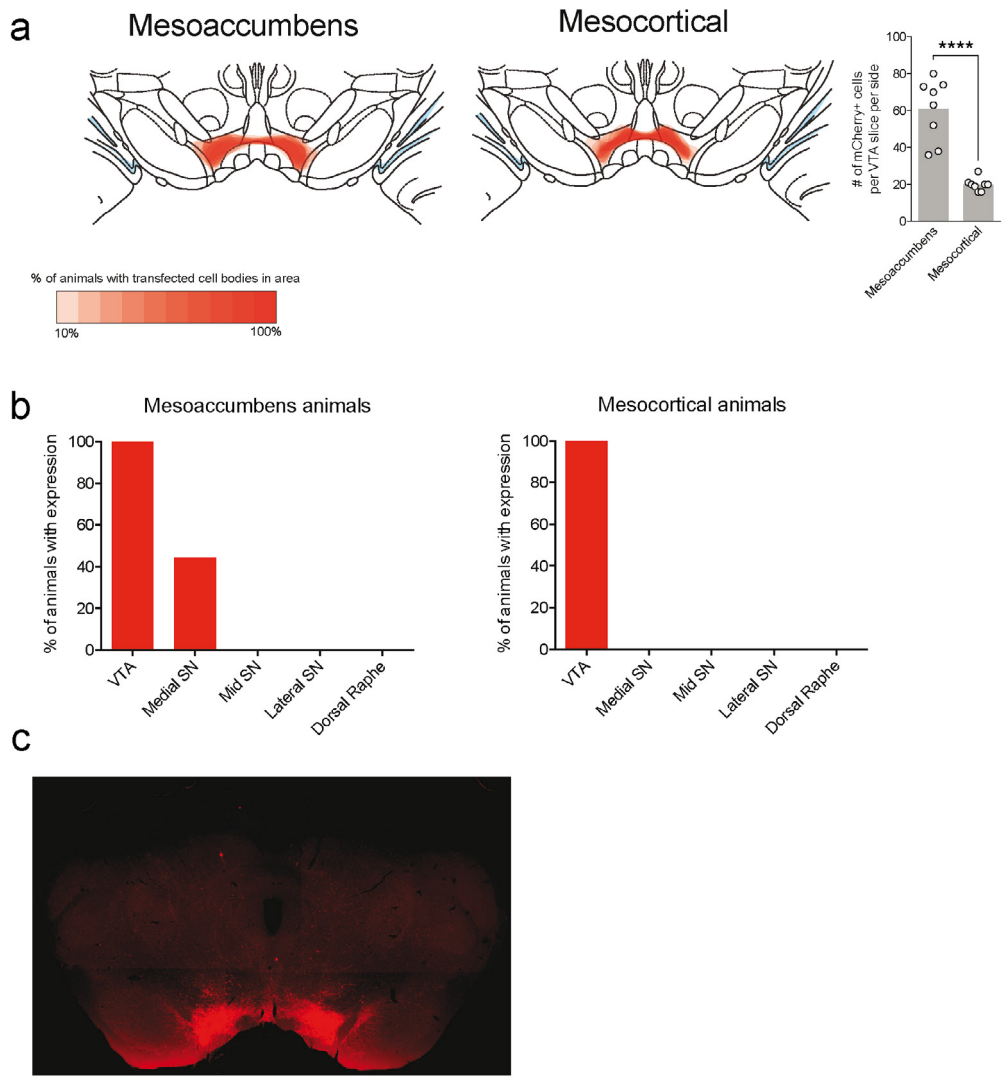


(a) Additional measures of the reversal learning task
(left panel) Plot of the cumulative reversals over time for all animals after systemic drug injection, confirming that the drug-induced performance impairment does not develop until after the first reversal (dashed line). Sidak's multiple comparisons test: $p < 0.05$ after trial 58 for D-amphetamine, $p < 0.05$ after trial 62 for cocaine.
(right panels)
Trials rewarded: one-way repeated measures ANOVA, $F(1.670, 40.08) = 3.998$, $p = 0.0327$. Post-hoc

(continuation of previous page)

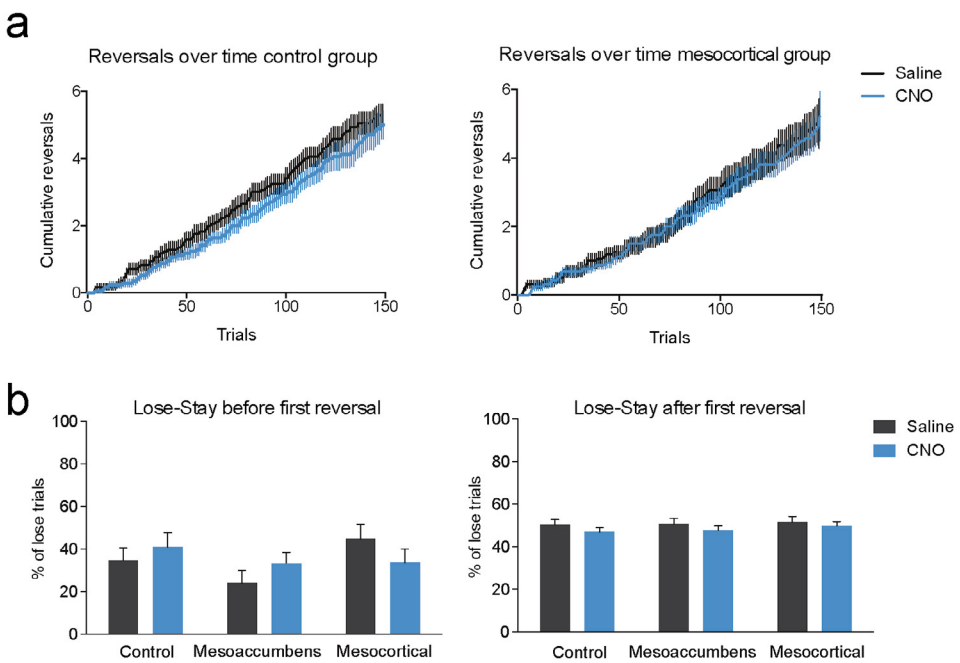
Sidak's test: cocaine vs saline, t(24) = 2.358, p = 0.0530; D-amphetamine vs saline, t(25) = 1.561, p = 0.2461.
Time to complete session: one-way repeated measures ANOVA, F(1.930, 46.33) = 3.454, p = 0.0415. Post-hoc Sidak's test: cocaine vs saline, t(24) = 2.388, p = 0.0497; D-amphetamine vs saline, t(25) = 0.2042, p = 0.9744.
Data shows mean ± standard error of the mean.
(b) Likelihood of model fits. Every dot represents an individual session. Cocaine and D-amphetamine did not significantly affect the fit of the model to the data (one-way repeated measures ANOVA, F(2, 48) = 1.783, p = 0.1791).
(c) Heatplot of simulated data showing how win- and lose-stay behavior (taken over the entire session) vary as a function of learning rates α_{win} and α_{loss}. Data shown are the average of 100 simulations of each $\alpha_{win}/\alpha_{loss}$ combination, with choice stochasticity factor β fixed at its mean for visualization purposes (β = 6.7). Dashed black lines show the average estimated learning rates after saline injection. The win-stay parameter is relatively stable for high learning rates compared to lose-stay, while lose-stay is more stable for lower learning rates. Hence, a decline of the average negative learning rate α_{loss} by ~2/3 more strongly affects win-stay than lose-stay behavior, providing an explanation for the observation that cocaine and D-amphetamine affect win-stay, but not lose-stay behavior. In contrast, when baseline learning rates would have been high, a decrease in α_{loss} would have resulted in an increase in lose-stay, without affecting win-stay behavior. Thus, how learning rates affect win- and lose-stay behavior is dynamic, and this strongly depends on the baseline estimates of α_{win}, α_{loss} and β.

SUPPLEMENTARY FIGURE 2



(a) (left) Spread of expression of Gq-mCherry in the midbrain. Shown is -5.40 mm posterior to Bregma. Atlas image adapted from Supplementary reference 1.
(right) Quantification of number of Gq-mCherry transfected neurons per group. Each dot represents a single animal. Significantly fewer neurons were transfected in the mesocortical group compared to the mesoaccumbens group (unpaired t-test, t(14) = 6.713, p < 0.0001).
(b) Quantification of expression of Gq-mCherry in the midbrain. In mesoaccumbens animals, virus sometimes spread to the medialmost part of the substantia nigra (SN), although this was always less than 5% of total transfected neurons.
(c) Example histology image of an animal from the mesoaccumbens group, showing strong expression of Gq-mCherry in the VTA and modest expression in the medial SN.

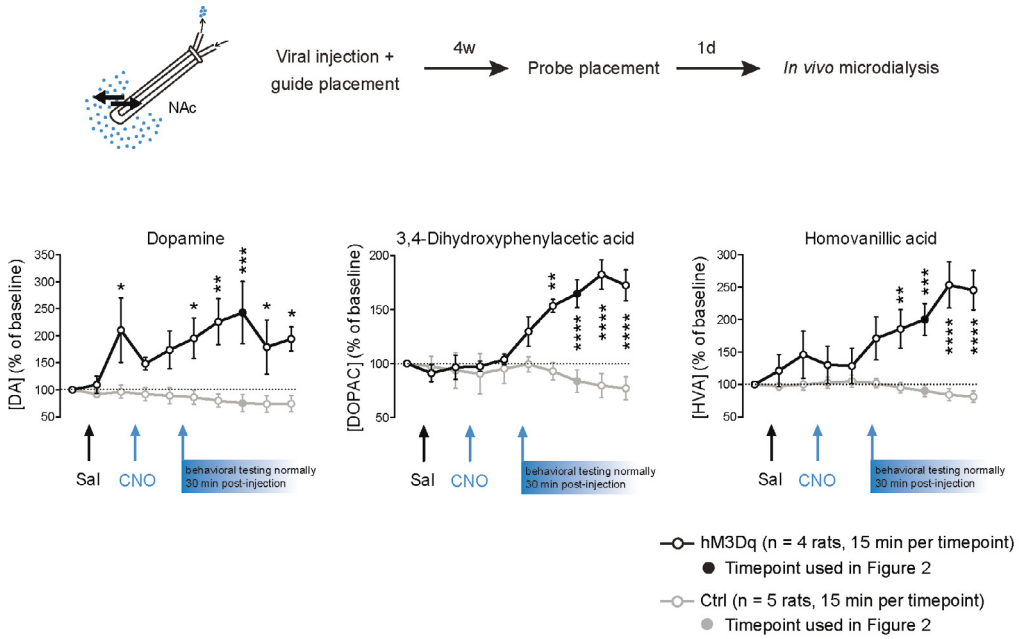
SUPPLEMENTARY FIGURE 3



(a) No effect of CNO treatment on the cumulative reversals over time for the control group and the mesocortical group (two-way repeated measures ANOVA for control group: main effect of CNO, $F(1, 16) = 2.919$, $p = 0.1068$; trials × CNO interaction, $F(149, 2384) = 0.7633$, $p = 0.9838$; two-way repeated measures ANOVA for data mesocortical group: main effect of CNO, $F(1, 15) = 0.2858$, $p = 0.6007$; trials × CNO interaction, $F(148, 2220) = 0.5058$, $p > 0.9999$).
(b) Lose-stay behavior during reversal learning is not affected by DREADD stimulation of either pathway.
Left: two-way repeated measures ANOVA; main effect of CNO, $F(1, 40) = 0.1325$, $p = 0.7178$; group × CNO interaction, $F(2, 40) = 2.136$, $p = 0.1314$.
Right: two-way repeated measures ANOVA; main effect of CNO, $F(1, 50) = 1.392$, $p = 0.2436$; group × CNO interaction, $F(2, 50) = 0.045$, $p = 0.9556$.

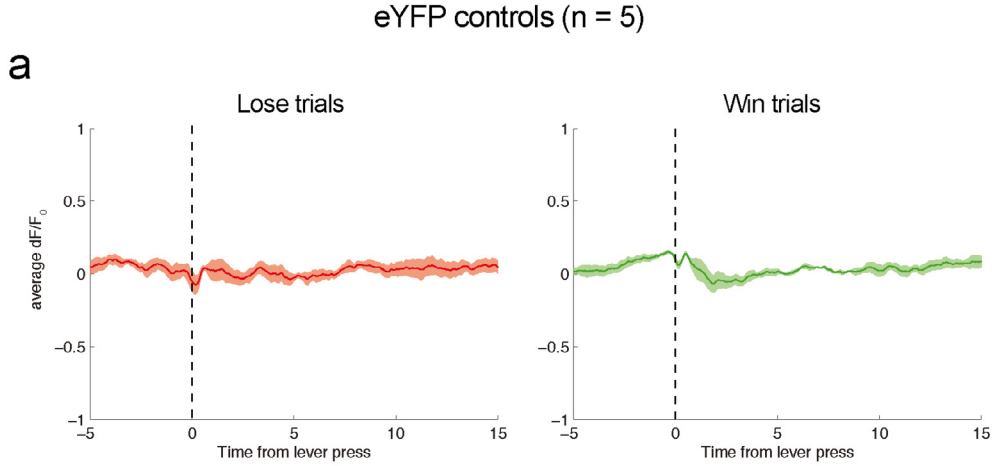
← (c) (left panel) Model fit on the reversal learning data of the mesoaccumbens group. DREADD activation altered α_{loss} in the same direction as cocaine and D-amphetamine, although not significantly so (one-tailed Wilcoxin matched-pairs signed rank test, $W = -41.00$, $p = 0.1764$). (right panel) Mesoaccumbens activation resulted in a significantly poorer fit of the model to the data (paired t-test, $t(16) = 3.224$, $p = 0.0053$). This seems consistent with the observation that during mesoaccumbens hyperactivity, both win-stay (Fig. 2e) and lose-stay behavior (Supplementary Figure 3b) are around chance level (50%), making the Rescorla-Wagner model a suboptimal descriptor of the animals' behavior.

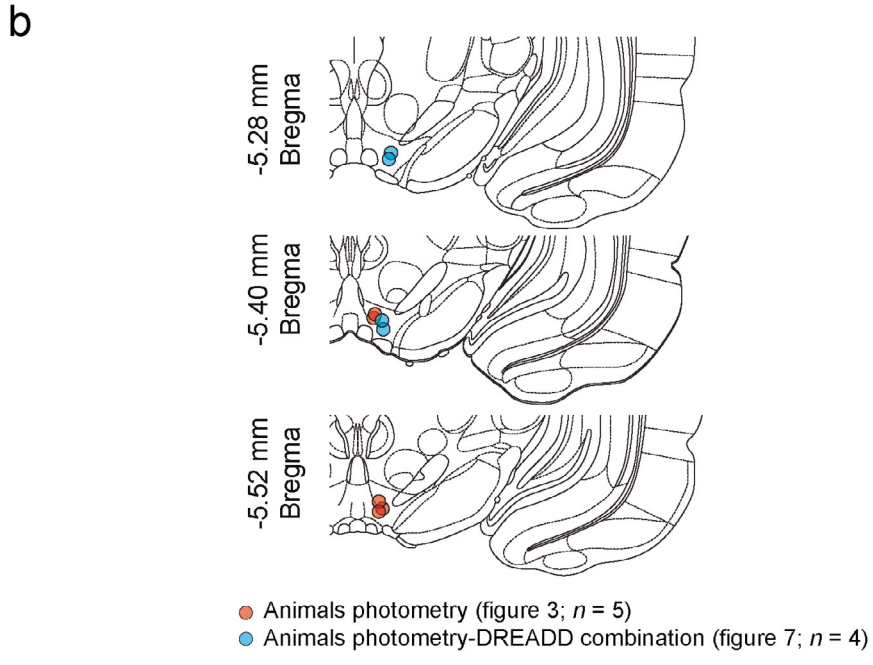
SUPPLEMENTARY FIGURE 4



In vivo microdialysis performed in the NAc showed increased baseline levels of DA and its metabolites after activation of the mesoaccumbens pathway by CNO (n = 4 animals DREADD group, n = 5 animals control group)
Two-way repeated measures ANOVA, with factors treatment and timepoints:
DA:
Main effect of treatment: $F(1,7) = 11.83$, $p = 0.0108$
Treatment × Time interaction effect: $F(9,63) = 4.11$, $p = 0.0003$
DOPAC:
Main effect of treatment: $F(1,7) = 9.77$, $p = 0.0167$
Treatment × Time interaction effect: $F(9,63) = 15.69$, $p < 0.0001$
HVA:
Main effect of treatment: $F(1,7) = 9.01$, $p = 0.0199$
Treatment × Time interaction effect: $F(9,63) = 23.65$, $p < 0.0001$
Post-hoc LSD tests: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Note a possible type I error at time point 3 in the DA graph.

SUPPLEMENTARY FIGURE 5





Fiber photometry
(a) Photometry responses during reversal learning in animals injected with the control fluorophore AAV-hSyn-eYFP (mean ± standard error of the mean).
(b) Fiber placement of animals used in photometry recordings. Atlas image adapted from Paxinos & Watson.

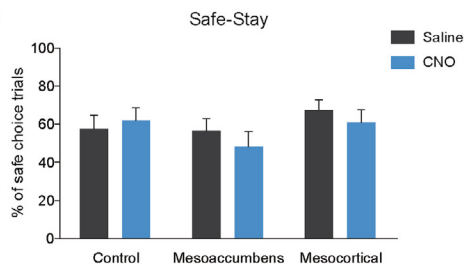
SUPPLEMENTARY FIGURE 6

a

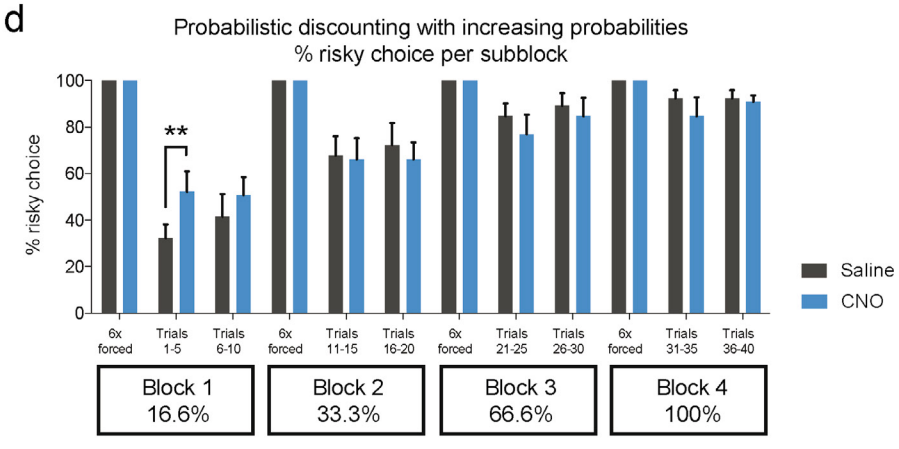
b

	Probabilistic discounting with decreasing probabilities	Probabilistic discounting with increasing probabilities
Block 1	Devaluation or revaluation of safe and risky levers, depending on expectations	Devaluation or revaluation of safe and risky levers, depending on expectations
Block 2	Devaluation of risky lever	Revaluation of risky lever
Block 3	Devaluation of risky lever	Revaluation of risky lever
Block 4	Devaluation of risky lever	Revaluation of risky lever

c

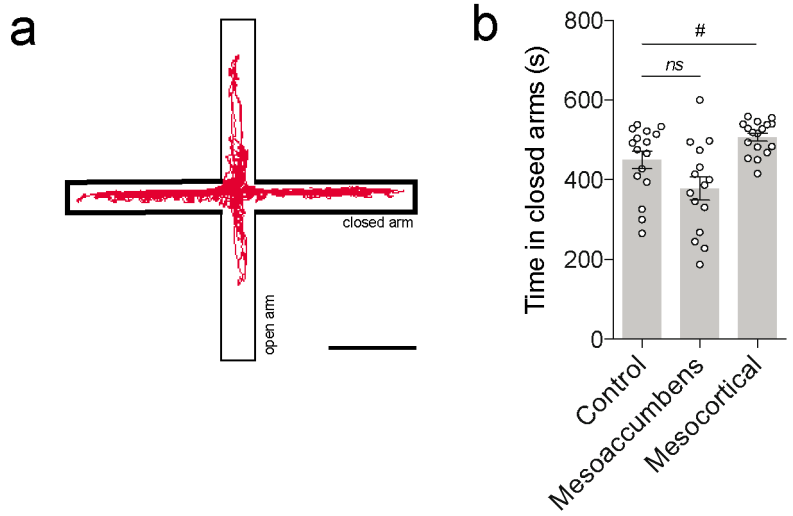


d



← Probabilistic discounting task.
(a) In the probabilistic discounting task with decreasing probabilities across trial blocks, responding on the risky lever is economically beneficial in the first block, responding on the safe lever is beneficial in the last two blocks. In the second block, the yield of both levers is equal. The opposite is true for the version of the task with increasing probabilities across trial blocks.
(b) Depending on a priori knowledge, in the first block of the probabilistic discounting task, de- and revaluative mechanisms are needed to determine the reward value of the safe and risky levers. Assuming that a proper neuronal representation of lever value has been established at the end of the first block, subsequent blocks in the probabilistic discounting task with decreasing probabilities (left column) involve devaluative mechanisms, whereas the probabilistic discounting task with increasing probabilities (right column) involve revaluative mechanisms.
(c) Safe-stay behavior, defined as the percentage of safe choice trials followed by another safe choice, was unaffected by CNO treatment (two-way repeated measures ANOVA, main effect of CNO, $F(1, 34) = 1.050$, $p = 0.3127$; group × CNO interaction, $F(2, 34) = 1.365$, $p = 0.2690$).
(d) Percentage choice of the risky lever in the probabilistic discounting task with increasing probabilities during mesoaccumbens stimulation. Only in the first 5 trials of block 1, mesoaccumbens activation increased the choice for the risky lever, despite the low chance on reward (Fisher's LSD test in block 1: $t = 2.652$, $p = 0.0096$. In all other blocks: $p > 0.2$).

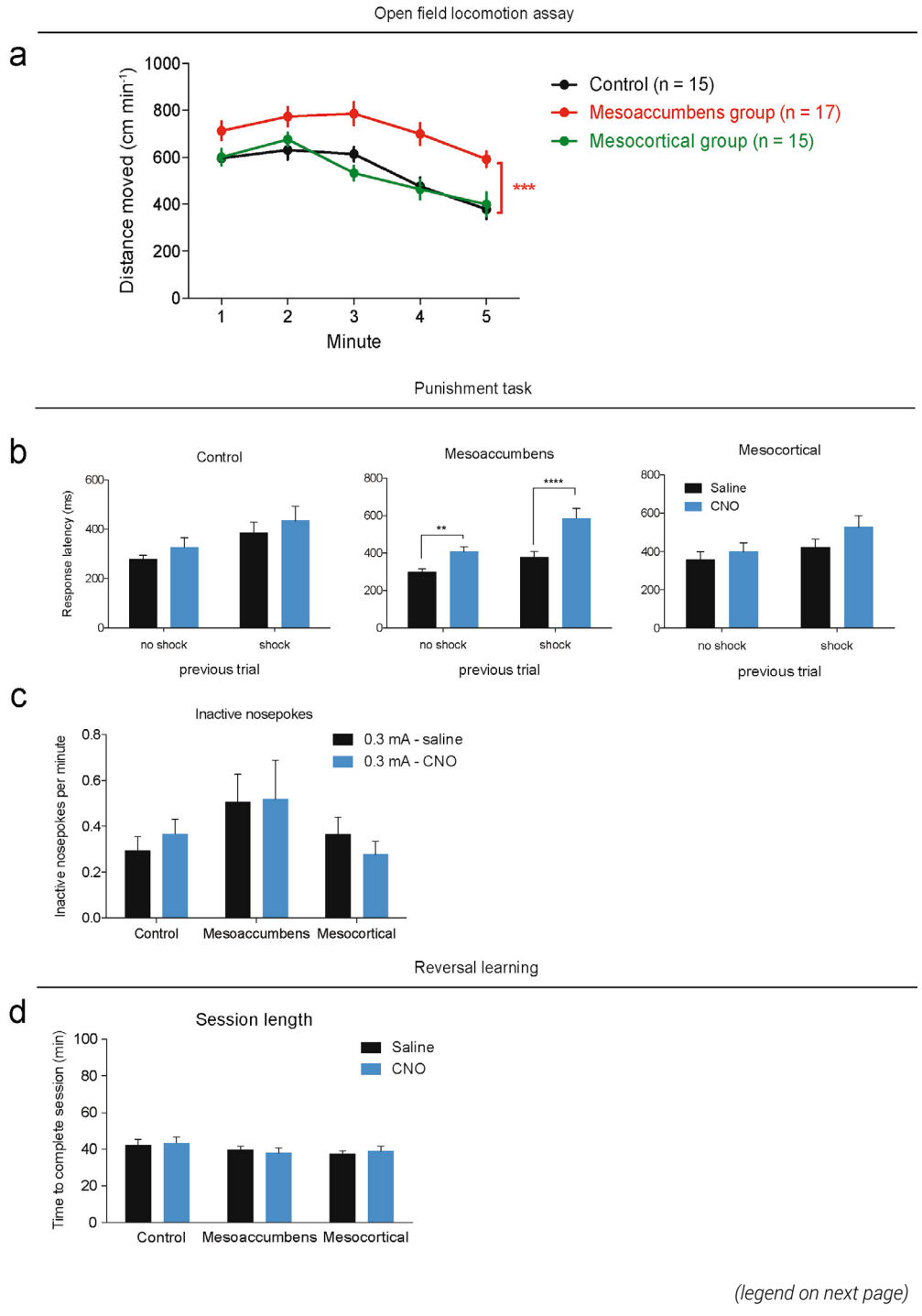
SUPPLEMENTARY FIGURE 7



Elevated plus maze

(a) Example track of a control animal in the elevated plus maze. Red line indicates the track of the animal's center point. Scalebar, 25 cm.
(b) Total time spent in the closed arms of the elevated plus maze. Stimulation of the mesocortical pathway showed a trend towards increased anxiety, whereas stimulation of the mesoaccumbens pathway had no effect on behavior (unpaired t-test with Welch's correction for unequal variance, Bonferonni corrected for 2 comparisons; $F(26.22)_{uncorrected} = 1.943$, $p = 0.1256$ for mesoaccumbens versus control, $F(21.25)_{uncorrected} = 2.378$, $^{\#}p = 0.053$ for mesocortical versus control). n = 16 control, n = 15 mesoaccumbens, n = 17 mesocortical.

SUPPLEMENTARY FIGURE 8

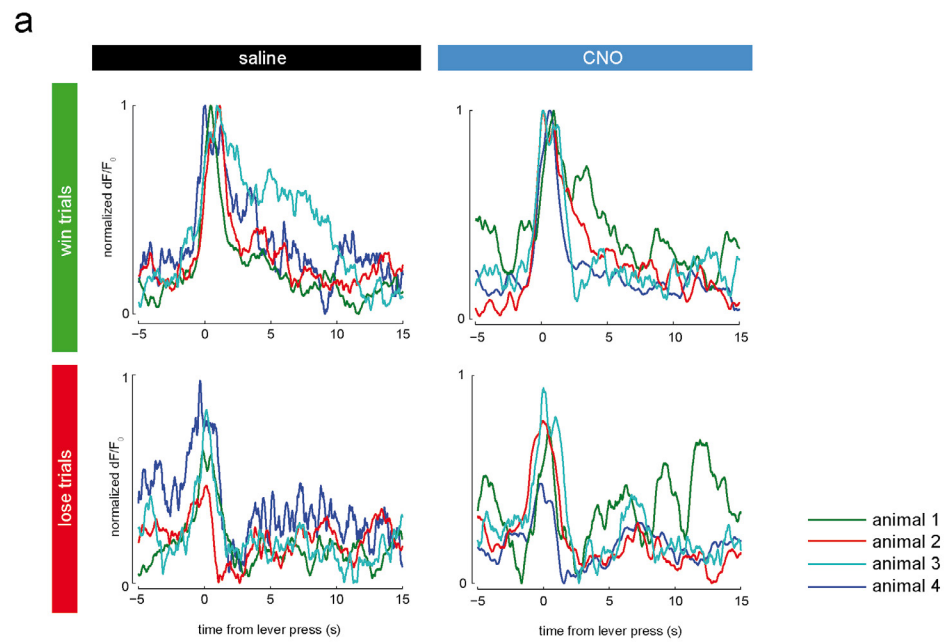


(legend on next page)

(continuation of previous page)

(a) Mesoaccumbens stimulation increases locomotion (Sidak's multiple comparisons test, mesoaccumbens versus control, $t(44) = 4.383$, $p = 0.0001$; mesocortical versus control, $t(44) = 0.1096$, $p = 0.9925$). All animals received CNO.
(b) Reaction times in the punishment task (based on trials 11-30). Receiving a foot shock during a trial robustly increased the reaction time during the subsequent trial in all three groups (two-way repeated measures ANOVA; main effect of shock, all groups $p < 0.01$). In addition, a significant main effect of CNO ($F(1,9) = 20.97$, $p = 0.0013$) and a significant shock × CNO interaction ($F(1,9) = 8.271$, $p = 0.0183$) were observed in the mesoaccumbens group. Post-hoc Sidak's multiple comparisons test revealed a significant slowing of responding after mesoaccumbens activation after a no-shock trial ($t(9) = 4.532$, $p = 0.0028$), as well as after a shock trial ($t(9) = 8.599$, $p < 0.0001$).
(c) Mesocortical or mesoaccumbens activation did not affect inactive nose poking in the punishment task (2-way repeated measures ANOVA; main effect of CNO, $F(1,25) < 0.0001$, $p = 0.9946$; group × CNO interaction, $F(2, 25) = 0.3164$, $p = 0.7316$).
(d) Time animals needed to complete the 150 trials of the reversal learning session was unaffected by CNO treatment (two-way repeated measures ANOVA; main effect of CNO, $F(1, 47) = 0.0439$, $p = 0.8350$; group × CNO interaction, $F(2, 47) = 0.2961$, $p = 0.7451$).

SUPPLEMENTARY FIGURE 9



Photometry recordings of VTA DA neurons during DREADD activation.

(a) Data from individual animals from figure 7a.
(b) To correct for bleaching, raw calcium signal was converted to dF/F_0 values by normalizing to a running-average baseline.

SUPPLEMENTARY TABLE 1

Model	# of free parameters	Parameter estimates (mean ± SEM) α_{win}	α_{loss}	β	aggregate LL	significance model improvement
M_0	0				-2599	
M_1	2	0.26 ± 0.05		2.0 ± 0.8	-2434	$M_1 > M_0$ $\chi^2(2) = 331.2$ $p = 0$
M_2	3	0.23 ± 0.06	0.31 ± 0.06	6.7 ± 1.7	-2421	$M_2 > M_1$ $\chi^2(1) = 24.9$ $p = 6.1 \times 10^{-7}$
	Constraints	[0 1]	[0 1]	[0 20]		

Model fits, performed on baseline behavior (i.e., after saline treatment) in the reversal learning task in the n = 25 rats from figure 1. Model 1 ('M_1') is the classical Rescorla-Wagner model, whereas model 2 ('M_2') uses separate learning rates for reward (α_{win}) and punishment (α_{loss}) learning. Since the tested models are nested (M_1 is a special case of M_2), model comparison was performed using the likelihood-ratio test. M_0 is the baseline model, in which choice behavior is random (p = 0.5 for every trial).

SUPPLEMENTARY TABLE 2

	Parameter estimates α_{win} Learning from positive RPE	α_{loss} Learning from negative RPE	β Choice stochasticity
Saline	0.23 ± 0.06	0.31 ± 0.06	6.7 ± 1.7
Cocaine	0.30 ± 0.07	0.13 ± 0.05 **	5.2 ± 1.4
D-amphetamine	0.26 ± 0.08	0.11 ± 0.02 *	8.8 ± 1.8

Best-fit model parameters, estimated by maximizing the log likelihood for the model given the choice sequences in every session. Wilcoxon matched-pairs signed rank test with Bonferroni correction, α_{loss}: cocaine versus saline, p = 0.0046; D-amphetamine versus saline, p = 0.032.

CHAPTER 3

Regional specialization of value-based learning functions in the rat prefrontal cortex

Jeroen P.H. Verharen
Hanneke E.M. den Ouden
Roger A.H. Adan*
Louk J.M.J. Vanderschuren*

* Equal contribution

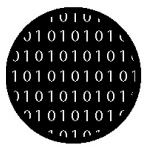
Manuscript in preparation

Highlights

- We pharmacologically inactivated different regions of the rat PFC during a probabilistic reversal learning task
- Computational modeling analyses revealed a whole-PFC involvement in punishment learning
- Additionally, prelimbic cortex and lateral OFC are involved in reward learning, and infralimbic cortex and medial OFC in choice perseveration

Techniques

Behavioral pharmacology



Computational modeling

CHAPTER 3

Value-based learning is a fundamental cognitive process that enables an organism to flexibly adapt to a changeable environment. To study how the rodent prefrontal cortex (PFC) contributes to this process, we assessed the effects of pharmacological inactivation of four PFC subregions on performance in a probabilistic reversal learning task in rats. Computational trial-by-trial analysis of the behavioral data revealed robust, whole-PFC coding of negative feedback learning. In contrast, positive feedback learning depended on function of the prelimbic and lateral orbital PFC, whereas response persistence required functional activity within the infralimbic and medial orbital PFC. As a result, pharmacological inactivation of any of the four subregions impaired reversal learning performance, either by reducing the number of reversals achieved (infralimbic, lateral orbital PFC) or rewards obtained (prelimbic and medial orbital PFC). In sum, our data show that distinct components of value-based learning are generated in medial and orbital PFC regions, displaying functional specialization and overlap. This organization suggests an intricate balance between efficiency and safeguarding of function within the rodent PFC.

To be able to survive and thrive in a dynamic environment, organisms must learn to repeat actions that were profitable in the past, while withholding actions that were not. For example, when a certain action leads to food reward, a hungry animal is likely to repeat that action. Conversely, when an action does not result in reward, or when it results in explicit punishment, an animal is likely to avoid that action in the future. This integration of action-outcome relationships is the basis of reinforcement learning theory[1-3], which states that value is attributed to preceding actions, updated based on their outcomes, and cached for when confronted with a similar choice later on. Such learning processes enables a system to flexibly adapt to a changing world and use environmental resources optimally.

It has long been known that function of the prefrontal cortex (PFC) underlies these value-based learning and decision making processes[4-8]. For example, lesions of different parts of the rodent PFC impair value-based decision making tasks like reversal learning[9,10], set shifting[11] and probabilistic discounting[12]. Importantly, value-based decisions are the result of a dynamic process weighing outcome expectancies, innate preferences and explorative urges, and alterations in overt behavior do not necessarily inform about which component processes are changed. Here, we sought to study the anatomical organization of these core neurocomputational processes underlying value-based learning and decision making in the rat PFC.

To this aim, we tested animals in a probabilistic reversal learning task[9,13] after pharmacological inactivation of four major subregions of the PFC that have been implicated in different aspects of value-based behavior[4-12]; the prelimbic cortex (PrL), the infralimbic cortex (IL), the medial orbitofrontal cortex (mOFC) and the lateral orbitofrontal cortex (lOFC). In the task, animals could earn sucrose pellets by responding into one of two holes that differed in the probability of delivering reward; one high-probability hole that gave 80% chance of reward and one low-probability hole that gave 20% chance of reward (Fig. 1a,b). Depending on the performance of the animals, reward contingencies switched between the two response options throughout the session, so that animals had to track the value of the options by integrating past wins and losses into a net expected value. After the animals had reached stable performance (after ~10 training sessions, see Online Methods), we infused a cocktail of the GABA receptor agonists baclofen and muscimol (or saline) into one of the four subregions of the PFC (Fig. S1).

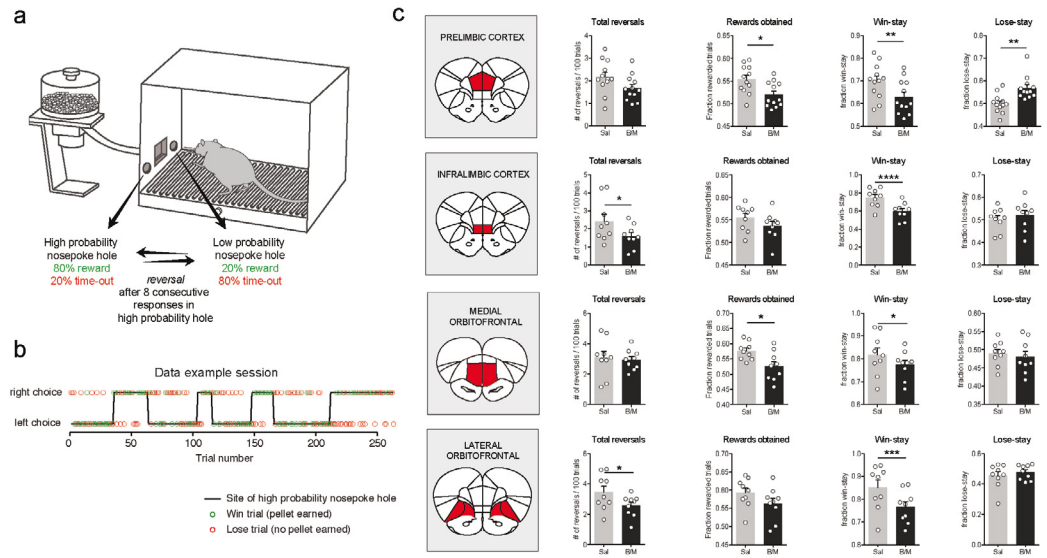
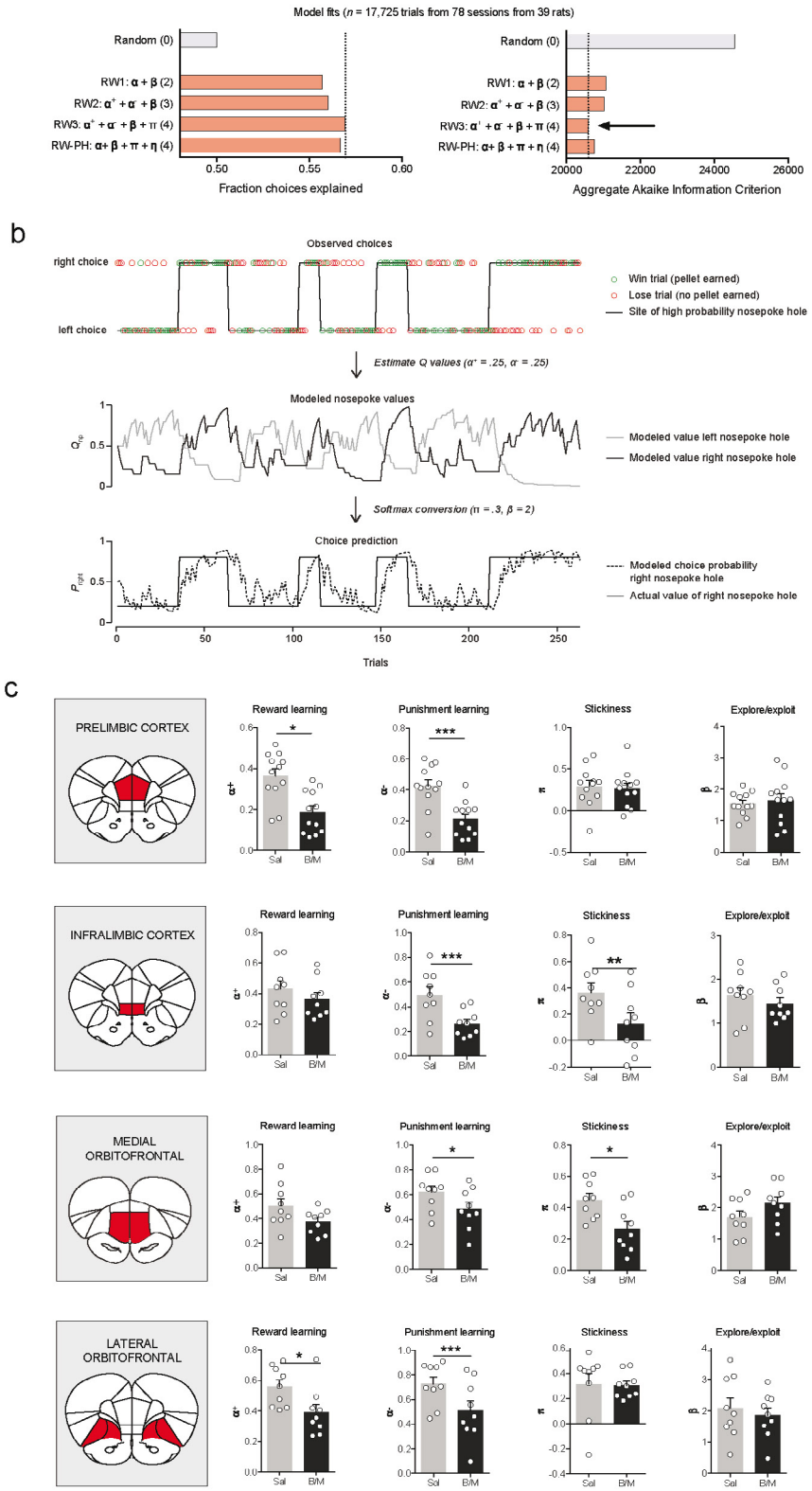


Figure 1 PFC inactivation impairs probabilistic reversal learning
a. Probabilistic reversal learning setup
b. Example session of one rat
c. Pharmacological PFC inactivation impaired task performance. See Supplementary statistics table for statistics; * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001 (post-hoc Holm-Sidak test).

Inactivation of any of the four PFC regions affected probabilistic reversal learning, but in different ways (Fig. 1c and Fig. S2). Inactivation of the PrL and mOFC reduced the total number of rewards the animals obtained, indicative of impaired performance. IL and lOFC inactivation did not change this outcome measure, but did result in a significant reduction in the total number of reversals the animals achieved (i.e., the rats less often reached the criterion of 8 consecutive responses at the high-probability nosepoke hole). Despite the fact that this did not lead to explicit negative consequences for the animals, a reduction in the total number of reversals indicates lower task volatility (as reward contingencies switched less often), which may be easier for the animals and may therefore mask a reduction in performance. Further analyses of the data demonstrated that inactivation of all regions reduced win-stay behavior, and PrL inactivation also increased lose-stay behavior. These win- and lose-stay measures provide a quantitative explanation for the behavioral response to the outcome of the last trial. However, they are poor indicators of positive and negative feedback learning when the animals more gradually switch behavior in response to changes in reward contingencies[14], not least because in the task used here, the animals have to track the value of a response over an extended history of trial outcomes, rather than just one.

To gain insight into the computational mechanisms subserving reversal learning that may be disrupted by the pharmacological inactivations, we fit a series of Q-learning models to the data. These models assume that the animals perform the task in order to maximize reward, by using past outcomes to track the value of each of the two nosepoke holes, and make choices based on these cached values. The first model we tested is the classic Rescorla-Wagner Q-learning model, where the value of each choice option is updated



according to the prediction error[1], i.e., the difference between the expected outcome and the actually received outcome according to learning rate α. Considering that a wealth of literature implicates the PFC in learning and decision making[15], this model should be able to demonstrate any general deficits in value-based learning caused by the inactivations. We next extended this model in various ways. Model 2 included separate learning rates for negative and positive feedback, α^+ and α^-, to allow for a certain manipulation to impact only one type of feedback learning, for example reward learning but not punishment learning. In model 3, we added a stickiness parameter π to this second model to assess the degree to which an animal perseverates on one choice option, independent of prior outcomes[16]. Model 4 was a Rescorla-Wagner/Pearce-Hall hybrid model[17,18] which was used to assess whether the learning rate changes when task volatility is higher (i.e., in proportion to the absolute prediction error, for example after a reversal). For all models, the value estimates were converted to choice probabilities using a Softmax function, allowing choice behavior to be stochastic to an extent described by parameter $1/\beta$ (often called the explore/exploit parameter; see online Methods). To estimate which of these learning mechanisms best described the animals' behavior, we fit these models to each individual reversal learning session and performed random effects model selection across all the baseline sessions, using the individual log-model-evidence estimates[19] (Fig. 2a).

Model 3 provided the best fit to the data (protected exceedance probability = 1; see Supplementary table 1), and explains the behavior of the animal on the basis of reward and punishment learning rates α^+ and α^-, stickiness parameter π, and stochasticity parameter β (Fig. 2b). Assessing the parameter values as a function of inactivation condition revealed differential contributions of the four PFC subregions to these different computational building blocks of value-based decision making (Fig. 2c). Pharmacological inactivation of the PrL and lOFC decreased both positive and negative learning rates, indicative of a reduced integration of past outcomes into future decisions. In contrast, IL and mOFC inactivation impaired negative, but not positive, learning rates, and also decreased stickiness, revealing that these animals showed less persistence on the same choice option; note that this is not necessarily disadvantageous. Importantly, estimates of stochasticity parameter β were unchanged across the inactivations, suggesting that pharmacological inactivation of the PFC affected value-based *learning*, but not value-based *decision making*. Although the different PFC regions show overlap in function (Fig. 3), inactivation did not always evoke the same changes in the classic measures of task performance (Fig. 1c), which is likely the result of an interaction between differences in baseline between the groups and differences in the strength of the effects of the PFC inactivations on the computational model parameters.

Given the observed overlap in value-based learning function, we speculate that in the rat PFC, there is (1) redundant coding of value-based learning and/or (2) coding of value-based learning function across a larger, interconnected network, that eventually mediates decision making. The former option could be indicative of the existence of a neural safety net, that ensures that essential cognitive operations can continue if activity in a part of the

Figure 2 Q-learning model coefficients
a. We fit several reinforcement learning models to our data, and estimated which model (i.e., strategy) best described the animals' behavior. Numbers in parentheses refer to the number of free parameters in the model.
b. The "winning" model was a Rescorla Wagner model, in which the animals dynamically track the value of both nosepokes by learning from reward and (omission) punishment.
c. Best-fit model parameters for each session. Inactivation of the PrL and lOFC impaired reward and punishment learning, whereas inactivation of the IL and mOFC impaired punishment learning and reduced choice perseveration (i.e. repeated choices for the same nosepoke hole). * P < 0.05, ** P < 0.01, *** P < 0.001 (post-hoc Holm-Sidak test).

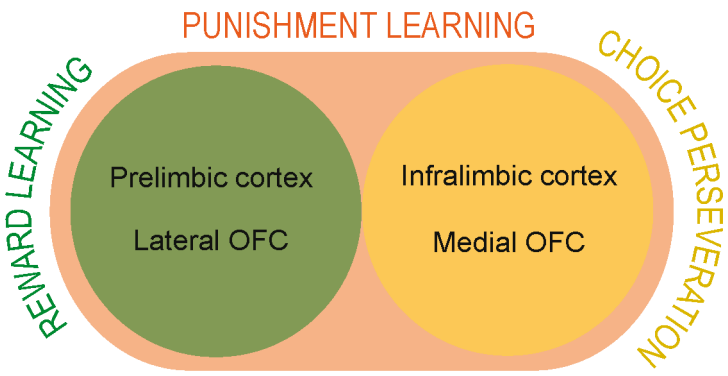


Figure 3 Visual summary. The four studied PFC regions have distinct, albeit overlapping functions in value-based decision making. All regions are involved in punishment learning.

PFC is impaired, for example by neurological disease, pharmacological insults, or stress. Alternatively, it may suggest distributed, parallel processing of value-related information across different brain circuits, as has been proposed by recent theories[20-22]. The latter option would implicate modular processing of information, so that dysfunction of one module of the network would directly hamper processing of feedback, leading to impairments in learning.

Deficits in reversal learning have been observed before after pharmacological inactivation or lesion of regions of the PFC across different species, although effects of neuronal manipulations of the medial PFC (PrL/IL) have been inconsistent[10]. It has been suggested that the medial PFC gets engaged in reversal learning only when the task becomes more complex and requires high attention[10], for example when reward contingencies become probabilistic. Indeed, most studies that have assessed the role of the medial PFC in reversal learning have used a deterministic version of the task, i.e., in which reward contingencies are non-probabilistic. Animals may then rely on more heuristic strategies to perform the task, such as win-stay/lose-switch[23], rather than by actively tracking the outcome of the choice options, thereby not requiring enhanced value-based learning function. One study used an experimental design almost identical to the one used in this study to assess the role of the rat PFC in reversal learning, and found similar changes in the classic measures of task performance after inactivation of the mOFC and lOFC[9]. However, they observed no effects of IL inactivation and a paradoxical increase in performance after PrL inactivation, which may be the result of our infusion being placed more dorsal (PrL) or anterior (IL) than theirs. Alternatively, it may be that the same neurocomputational processes were affected by the inactivation in their study, but that this did not lead to significant changes in the classic (compound) measures of task performance, perhaps because of differences in baseline performance. Indeed, the authors suggested that the seeming improvement in reversal learning performance after PrL inactivation may actually be the result of impairments in monitoring positive and negative outcomes of the task[9].

Previous work has indicated competing function of the PrL and IL in motivated behavior, with the PrL mediating goal-direct learning and the IL mediating habitual responding[24]. Our observation that both these areas are involved in negative feedback learning provides evidence against this antagonistic hypothesis, although the observed involvement of the PrL in positive feedback learning versus the IL in choice perseveration

may underlie the findings supporting this original hypothesis, as reward seeking and motor perseveration may logically be related to goal-directed and habitual behavior, respectively.

The OFC comprises a large part of the PFC and has been shown to be involved in a variety of decision making tasks, with functional heterogeneity along both the mediolateral and the anteroposterior axis[25]. An extensive list of functions have been ascribed to the OFC[26], most of which encompass some form of multisensory integration of reward and environment. In this study, we mainly targeted the anterior MO/VO region of the mOFC and the more dorsal part of the VO/LO region of the lOFC, which have been linked to decision making under uncertainty and outcome prediction, respectively[25]. However, lesions of the lOFC affect a larger array of tasks than just those involving outcome prediction, which suggests that lOFC function may be captured better by other theories, such as the more recent theory that the lOFC keeps a cognitive map of task structure, thereby serving different functions depending on the type of behavioral task[27]. One interesting hypothesis is that the mOFC may serve as a bridge between the lOFC and medial parts of the PFC (including PrL and IL)[25,28], suggesting serial processing of information, which may in part explain the shared functionality that we observed.

Overall, our study reveals a rat PFC that is anatomically organized into functional districts, in which each function underlying value-based learning depends on activity in at least two different PFC subregions. Such a topographic map of PFC function suggests an intricate balance between an efficient distribution of function, so that not all regions are engaged in all aspects of task behavior, and safeguarding of function, so that each function relies on activity in at least two brain regions. Interestingly, punishment learning was dependent on each of the four PFC regions, suggesting that negative feedback learning is especially robustly integrated in the frontal lobe, perhaps because of its importance for survival. Altogether, we demonstrate a specialized but overlapping functional-anatomical organization of higher-order cognition within the rat PFC, providing exciting new insights into the functional architecture of the mammalian brain.

Methods

1 Animals

48 adult male (>300 g) Long-Evans rats (Janvier labs, France) were used for the experiments. Rats were solitarily housed in a humidity- and temperature-controlled room and were kept on a 12h/12h reversed day/night cycle (lights off at 8AM). All experiments took place in the dark phase of the cycle. Animals were kept on food restriction (~5g standard lab chow per 100g body weight per day) during behavioral training and the experiments. All experiments were conducted in accordance with European (2010/63/EU) and Dutch (Wet op de Dierproeven, revised 2014) law, and approved by the Dutch Central Animal Testing Committee, and by the Animal Ethics Committee and the Animal Welfare Body of Utrecht University.

2 Surgeries

Animals were implanted with bilateral guide cannulas above each of the target areas (one brain area per group). For surgery, animals were anaesthetized with an i.m. injection of a mixture of 10 mg/kg fluanisone and 0.315 mg/kg fentanyl (Hypnorm, Janssen Pharmaceutica, Beerse, Belgium). Animals were placed in a stereotaxic apparatus (David Kopf Instruments, Tujunga, USA), and an incision was made along the midline of the skull. Using a dental drill, two small craniotomies were made above the area of interest, and 26G guide cannulas (Plastic Ones, Roanoke, USA) were lowered to the following positions (relative to Bregma):

PrL	AP +3.2 mm	ML ±0.6 mm	DV -2.6 mm from skull
IL	AP +3.2 mm	ML ±0.6 mm	DV -4.3 mm from skull
mOFC	AP +4.4 mm	ML ±0.6 mm	DV -3.8 mm from skull

lOFC AP +3.6 mm ML ±2.6 mm DV -3.7 mm from skull under a 5° angle

For the PrL, IL and mOFC groups, guide cannulas were used with a bilateral protrusion of 5 mm (with 1.2 mm space between the protrusions). For the lOFC group, single cannulas were used with a protrusion length of 5 mm.

Guide cannulas were secured with screws, dental glue (C&B Metabond, Parkell Prod Inc., Edgewood, USA) and dental cement, and the skin of the animals was sutured so that no skull was exposed. After the surgery, animals received saline (10 ml once, s.c.) and carprofen for pain relief (5 mg/kg, 3x daily, s.c.). Dummy injectors were placed into the cannula. Animals were allowed to recover for at least 7 days before behavioral training started.

3 Behavioral task

The behavioral task was conducted in operant chambers (Med Associates Inc., USA, 30.5×24.2×21.0 cm), placed within sound-attenuated cubicles. The boxes contained two illuminated nosepoke holes, a tone generator and a house light on one side of the chamber, and on the other side of the chamber a food receptacle delivering 45mg sucrose pellets (SP; 5UTL, TestDiet, USA) flanked by two cue lights.

At task initiation, one of the two nosepoke holes was randomly assigned as the high-probability hole, that gave 80% chance on reward and 20% chance on a time-out, and the other hole was assigned as the low-probability hole, which gave 20% chance on reward and 80% chance on a time-out (Fig. 1a). Determination of the response outcome (reward or time-out) happened through independent sampling, so that the outcome of the previous trial did not affect the odds of reward in the next trial. The start of the session was signaled to the animal by illumination of the house light and the two nosepoke holes.

Directly after a 'win' response (i.e., a responses in one of the two nosepoke holes that resulted in reward delivery), the lights in the two nosepoke holes were turned off, a sucrose pellet was delivered into the food receptacle, a tone was played for 0.5s, and the two cue lights next to the food receptacle were turned on. Consumption of the reward was measured by an infrared light sensor in the food receptacle, after which the cue lights were extinguished and a new trial was initiated. After a 'lose' response (i.e., a response in one of the two nosepoke holes that resulted in a time-out), the house light and lights in the nosepoke holes were turned off, and a 10s time-out started during which animals remained in the dark, and poking either of the two nosepoke holes was without scheduled consequences. After 10s, a new trial was automatically initiated, signaled to the animal by the illumination of the house light and the two nosepoke holes.

When the animal made 8 consecutive responses at the high-probability nosepoke hole, the contingencies reversed, so that the previously high-probability nosepoke hole became the low-probability nosepoke hole, and vice versa. The task automatically terminated after 90 minutes, and animals were allowed to make an unrestricted number of trials during this period.

The task was optimized for computational modeling by making two major changes to the classic probabilistic reversal learning paradigm. First, animals were allowed to make an unrestricted amount of trials during the 90-minute session, as there is a strong positive relation between reliability of model parameter estimation and the amount of trials on which that estimation is based. Second, there was no restriction to the time in which the animals could make a response at one of the nosepoke holes (i.e., no trials were designated as 'omissions'), because it is unknown how an omitted trial affects the value representation of the two nosepoke holes.

For each trial, the choice of the animal, the side of the high-probability nosepoke hole, the outcome of the trial (win or lose), and the timestamps of trial start and nosepoke response were monitored. Win-stay was defined as the fraction of win trials on which the animal chose that same nosepoke hole on the next trial. Lose-stay was defined as the fraction of lose trials on which the animal chose that same nosepoke hole on the next trial.

4 Pharmacological inactivations

Infusions took place when animals reached stable performance in the task, which was typically after ~10 training sessions, which was defined as a non-significant repeated measures one-way ANOVA on the total number of reversals per 100 trials for 3 consecutive days. One day before test sessions, all animals received an infusion of saline, to habituate them to the infusion procedure. The next days, animals received an infusion with saline or a cocktail of baclofen (1 nmol; Sigma-Aldrich, The Netherlands) and muscimol (0.1 nmol; Sigma-Aldrich, The Netherlands) dissolved in saline, counterbalanced between days, with 24h in between.

For the infusion, dummy injectors were removed and replaced by injectors that injected 0.3 µl/side of the dissolved drug (or saline) at a rate of 1 µl/min with a syringe pump (Harvard apparatus, Holliston, USA). The injectors were kept in place for an additional 30 seconds after the infusion to allow for proper diffusion of the drug into the tissue. Injectors of the double cannulas protruded 1 mm, and the injectors of the single cannulas protruded 0.4 mm below the termination point of the guide cannula. After the infusion, the animals were placed back in their home cage for 10 minutes, after which they were placed in the operant boxes.

To reduce intra-animal variability, thereby reducing the number of animals necessary to achieve the same statistical power, we repeated the experiment a second time, and averaged all task measures across the two conditions. In other words, animals were measured twice after saline infusion, and twice after baclofen+muscimol infusion, and the outcomes were averaged to get one single saline, and one single baclofen+muscimol measure, which was used in further analyses.

5 Computational modeling

5.1 Basic model

We fit a series of Q-learning models to our data to assess which model (i.e., strategy) best described the animals' behavior in the task. The first model we tested is the classic Rescorla-Wagner Q-learning model (RW1), that assumes that on every trial t, the nosepoke values are updated based on the reward prediction error (RPE), which is the difference between the reward received (this is 1 for win trials, 0 for lose trials) and the reward expected (i.e., the expected value Q of the chosen nosepoke hole s):

$$RPE_t = outcome_t - Q_{s,t-1} \quad (1)$$

so that

$$RPE_t = \begin{cases} 1 - Q_{s,t-1} & \text{for win trials} \\ 0 - Q_{s,t-1} & \text{for lose trials} \end{cases} \quad (2)$$

Nosepoke hole values were subsequently updated with learning rate α according to a Q-learning rule:

$$Q_{s,t} = Q_{s,t-1} + \alpha \times RPE_t \quad (3)$$

Note that the value of the unchosen side was not updated and thus retained its previous value. For the first trial, both nosepoke values were initiated at 0.5.

The relationship between nosepoke values Q_{left} and Q_{right}, and the probability that the rat chooses the left or right ($p_{left,t}$, respectively $p_{right,t}$) nosepoke hole in every trial was described by a Softmax function:

$$p_{right,t} = \frac{\exp(\beta \cdot Q_{right,t})}{\exp(\beta \cdot Q_{left,t}) + \exp(\beta \cdot Q_{right,t})} \quad (4)$$

and $p_{left,t} = 1 - p_{right,t}$ (5)

In this function, β is the Softmax inverse temperature, which indicates how value-driven the agent's choices are. If β becomes very large, then the value function $\beta \cdot Q_{s,t}$ of the highest valued side becomes dominant, and the probability that the agents chooses that side approaches 1. Is β zero, then $p_{left,t} = p_{right,t} = e^0/(e^0 + e^0) = 0.5$. β is sometimes referred to as the explore/exploit parameter, where a low β favors exploration (i.e., sampling of all options) and a high β favors exploitation (i.e., choosing the option which has proven to be beneficial). Therefore, a decrease in β may reflect a more general disruption of behavior, since it indicates that the animal chose more randomly.

All the subsequently tested models are extensions of this Rescorla-Wagner Q-learning model.

5.2 Model extensions

The second model we tested (RW2) is similar to RW1, except that separate learning rates were used for learning from positive (reward delivery; win trials) and negative (reward omission; lose trials) feedback, α^+ and α^-, respectively. The value updating function is thus given by equation 6:

$$Q_{s,t} = \begin{cases} Q_{s,t-1} + \alpha^+ \cdot \text{RPE}_t & \text{for win trials} \\ Q_{s,t-1} + \alpha^- \cdot \text{RPE}_t & \text{for lose trials} \end{cases} \quad (6)$$

Model RW3 is an extension of model RW2 and adds a stickiness parameter π to the model. This parameter indicates a preference for the previously chosen ($\pi > 0$; perseveration) or previously unchosen ($\pi < 0$; alternation) option, so that the Softmax is given by equation 7:

$$p_{right,t} = \frac{\exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})}{\exp(\beta \cdot Q_{left,t} + \pi \cdot \phi_{left,t}) + \exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})} \quad (7)$$

Here, ϕ is a boolean with $\phi = 1$ if that hole was chosen in the previous trial, and $\phi = 0$ if not. For example, if the right nosepoke hole was chosen in trial *t-1*, then $\phi_{right,t}$ becomes 1, and $\phi_{left,t}$ becomes 0. This adds a certain amount π to the value function of the right nosepoke hole in trial *t*, in addition to the nosepoke hole's expected value $Q_{right,t}$.

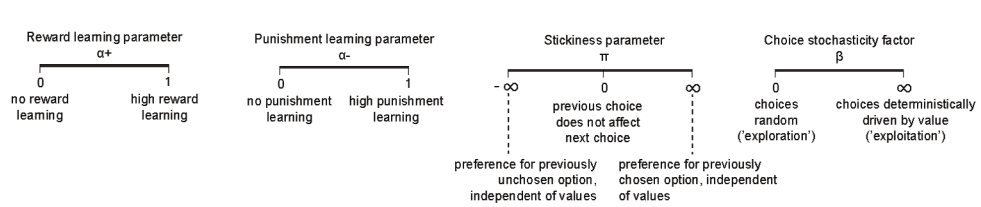
In addition, we tested a hybrid Rescorla-Wagner/Pearce-Hall model of reinforcement learning, that is able to account for an increased learning rate when task volatility is higher, for example right after a reversal. As such, it has a fixed single learning rate *α*, and a variable learning rate γ that is dependent on the unsigned prediction error to a certain amount η (which was a free variable in the model).

The following table shows the equations that are used for value updating and the conversion of values into action probabilities for each of the five models:

Model	Free parameters	Learning model	Observation equation
RW1	α, β	$Q_{s,t} = \begin{cases} Q_{s,t-1} + \alpha \cdot \text{RPE}_t & \text{for win trials} \\ Q_{s,t-1} + \alpha \cdot \text{RPE}_t & \text{for lose trials} \end{cases}$	$p_{right,t} = \frac{\exp(\beta \cdot Q_{right,t})}{\exp(\beta \cdot Q_{left,t}) + \exp(\beta \cdot Q_{right,t})}$
RW2	α+, α-, β	$Q_{s,t} = \begin{cases} Q_{s,t-1} + \alpha^+ \cdot \text{RPE}_t & \text{for win trials} \\ Q_{s,t-1} + \alpha^- \cdot \text{RPE}_t & \text{for lose trials} \end{cases}$	$p_{right,t} = \frac{\exp(\beta \cdot Q_{right,t})}{\exp(\beta \cdot Q_{left,t}) + \exp(\beta \cdot Q_{right,t})}$
RW3	α+, α-, β, π	$Q_{s,t} = \begin{cases} Q_{s,t-1} + \alpha^+ \cdot \text{RPE}_t & \text{for win trials} \\ Q_{s,t-1} + \alpha^- \cdot \text{RPE}_t & \text{for lose trials} \end{cases}$	$p_{right,t} = \frac{\exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})}{\exp(\beta \cdot Q_{left,t} + \pi \cdot \phi_{left,t}) + \exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})}$
RW-PH	α, β, π, η	$Q_{s,t} = \begin{cases} Q_{s,t-1} + \alpha \cdot \gamma_t \cdot \text{RPE}_t & \text{for win trials} \\ Q_{s,t-1} + \alpha \cdot \gamma_t \cdot \text{RPE}_t & \text{for lose trials} \end{cases}$ with $\gamma_t = \eta \cdot \lvert \text{RPE}_t \rvert + (1 - \eta) \cdot \gamma_{t-1}$	$p_{right,t} = \frac{\exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})}{\exp(\beta \cdot Q_{left,t} + \pi \cdot \phi_{left,t}) + \exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})}$

In this table, α = Rescorla-Wagner learning rate, β = choice stochasticity, π = stickiness factor, η = Pearce-Hall associability factor, $Q_{s,t}$ = value of nosepoke *s* on trial *t*, $p_{s,t}$ = choice probability of nosepoke *s* on trial *t*, ϕ = boolean that is 1 if nosepoke *s* is chosen on the previous trial and 0 if unchosen on previous trial, RPE = reward prediction error, and γ_t = associability on trial *t*.

An overview of the interpretation of the parameters of the 'winning' RW3 model:



5.3 Parameter estimation

To obtain realistic estimates of the model parameters on a population level, we used maximum a posteriori probability (MAP) estimation. This was done because a simple grid search sometimes lead to unrealistic parameter values (for example, learning rates >1). The used priors for the MAP estimation were:

α^+, α^-	betapdf(1.5, 1.5)
π	normpdf(0.5, 0.5)
β	normpdf(2, 2)

Multiplication of these priors with the likelihood gives the posterior probability of the model parameters given the observed choice sequence:

$$P(\{\alpha^+, \alpha^-, \pi, \beta\} \mid \text{data,model}) = P(\text{data} \mid \text{model}, \{\alpha^+, \alpha^-, \pi, \beta\}) \cdot P(\{\alpha^+, \alpha^-, \pi, \beta\} \mid \text{model}) \quad (8)$$

in which P(data | model, {α^+, α^-, π, β}) is the likelihood of the observed choice sequence (from trial 1 to the last trial *T*) given the model and the parameter settings (computed as the log likelihood):

$$\log(P(\text{data}\,|\,\text{model},\{\alpha^{+},\alpha^{-},\pi,\beta,\eta\})) = \sum_{t=1}^{T} \log(P(\text{choice}_{t}\,|\,Q_{left,t},Q_{right,t},\phi_{left,t},\phi_{right,t})) \quad (9)$$

The posterior probability was calculated for many combinations of parameters {α^{+}, α^{-}, π, β}, and arranged in a multidimensional grid. Best-fit parameter values were then estimated by integrating these posterior probabilities over the parameter's range, marginalized over the other parameters.

5.4 Model comparisons

The log-model evidences of individual sessions were penalized for model complexity by computing the Akaike Information Criterion and Bayesian Information Criterion:

AIC = 2*[# of free parameters in the model] - 2*log(likelihood) (10)

BIC = -2*log(likelihood) + [# of free parameters in the model]*log([# of trials]) (11)

As such, a lower value of the AIC and BIC reflects more evidence in favor of the model. In addition, figure 2a contains a random choice model, in which all choices had a probability of 0.5, hence the log likelihood for each session was computed as $\log(0.5^{\text{total trials}})$. To compare models, we entered the AIC's of all baseline sessions (i.e., after saline infusion) in a random effects Bayesian model comparison (implemented in SPM12) analysis to assess the evidence that one model is more likely than any of the others (see ref. 19).

6 Statistics

Statistical tests were performed with Prism 6 (GraphPad Software Inc.). For each measure, a 2-way repeated measures analysis of variance (ANOVA) was used, in which drug (saline versus baclofen+muscimol) was used as a within-subjects repeated measures factor, and treatment group (PrL, IL, mOFC and lOFC) as a between-subjects factor. When the ANOVA yielded a significant interaction effect, or a main effect of drug ($p < 0.05$), a post-hoc repeated measures Holm-Sidak test was used to test, for each group, whether there was a significant difference between the saline and baclofen+muscimol sessions. All statistics are presented in the supplementary statistics table. In all figures: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

7 Data availability

All behavioral data is openly available at http://www.github.com/jeroenphv/.

8 Exclusion criteria

All experimental groups started with 12 animals. The following animals were excluded from the experiment or analysis:
PrL group: none (final group *n* = 12 rats).
IL group: 2 rats died during the surgery, 1 rat was excluded due to misplacement of the cannulas (final group *n* = 9 rats).
mOFC group: 2 rats died during the surgery, 1 rat was excluded due to misplacement of the cannulas (final group *n* = 9 rats).
lOFC group: 1 rat died during the surgery, 2 rats were excluded due to misplacement of the cannulas (final group *n* = 9 rats).

References

1. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 2, 64-99 (1972).
2. Sutton, R. S. & Barto, A. G. Reinforcement learning: An introduction. (MIT press, 1998).
3. Dayan, P. & Daw, N. D. Decision theory, reinforcement learning, and the brain. Cogn. Affect Behav. Neurosci. 8, 429-453 (2008).
4. Miller, E. K. & Cohen, J. D. An Integrative Theory of Prefrontal Cortex Function. Annu. Rev. Neurosci. 24, 167-202 (2002).
5. Dalley, J. W., Cardinal, R. N. & Robbins, T. W. Prefrontal executive and cognitive functions in rodents: neural and neurochemical substrates. Neurosci. Biobehav. Rev. 28, 771-784 (2004).
6. Roberts, A. C. Primate orbitofrontal cortex and adaptive behaviour. Trends in cogn. sci. 10, 83-90 (2006).
7. Robbins, T. W. & Arnsten, A. F. The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. Annu. Rev. Neurosci. 32, 267-287 (2009).
8. Floresco, S. B. Prefrontal dopamine and behavioral flexibility: shifting from an "inverted-U" toward a family of functions. Front Neurosci. 7, 62 (2013).
9. Dalton, G. L., Wang, N. Y., Phillips, A. G. & Floresco, S. B. Multifaceted Contributions by Different Regions of the Orbitofrontal and Medial Prefrontal Cortex to Probabilistic Reversal Learning. J. of Neurosci. 36, 1996-2006 (2016).
10. Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H. & Holmes, A. The neural basis of reversal learning: An updated perspective. Neuroscience (2016).
11. Birrell, J. M. & Brown, V. J. Medial Frontal Cortex Mediates Perceptual Attentional Set Shifting in the Rat. J. of Neurosci. 20, 4320-4324 (2000).
12. St Onge, J. R. & Floresco, S. B. Prefrontal cortical contribution to risk-based decision making. Cereb. Cortex 20, 1816-1828 (2010).
13. Bari, A. et al. Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. Neuropsychopharmacology 35, 1290-1301 (2010).
14. Verharen, J. P. H. et al. A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. Nat. comm. 9 (2018).
15. Miller, E. K. & Cohen, J. D. An Integrative Theory of Prefrontal Cortex Function. Annu. Rev. Neurosci. 24, 167-202 (2001).
16. Gershman, S. J. Empirical priors for reinforcement learning models. J. of Math. Psychol. 71, 1-6 (2016).
17. Pearce, J. M. & Hall, G. A Model for Pavlovian Learning: Variations in the Effectiveness of Conditioned But Not of Unconditioned Stimuli. Psychol. Rev. 87, 532-552 (1980).
18. Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A. & Daw, N. D. Differential roles of human striatum and amygdala in associative learning. Nat. neurosci. 14, 1250-1252 (2011).
19. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies - revisited. Neuroimage 84, 971-985 (2014).
20. Cisek, P. Making decisions through a distributed consensus. Curr. opin. in neurobiol. 22, 927-936 (2012).
21. Rushworth, M. F., Kolling, N., Sallet, J. & Mars, R. B. Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Curr. opin. in neurobiol. 22, 946-955 (2012).
22. Hunt, L. T. & Hayden, B. Y. A distributed, hierarchical and recurrent framework for reward-based choice. Nat. Rev. Neurosci.18, 172 (2017).
23. Posch, M. Win–Stay, Lose–Shift Strategies for Repeated Games—Memory Length, Aspiration Levels and Noise. J. Theor. Biol. 198, 183-195 (1999).

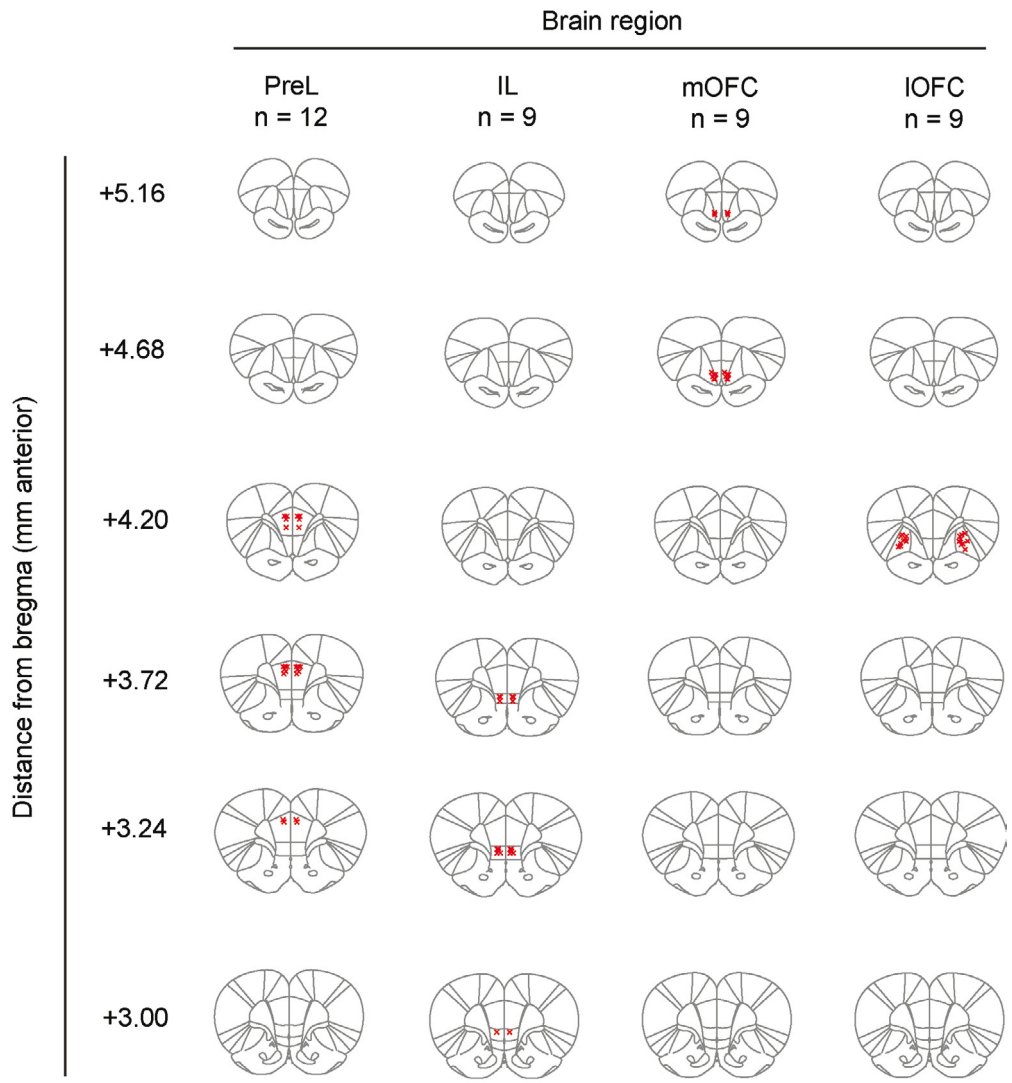
24. Balleine, B. W. & O'Doherty, J. P. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35, 48-69 (2010).
25. Izquierdo, A. Functional Heterogeneity within Rat Orbitofrontal Cortex in Reward Learning and Decision Making. J. of Neurosci. 37, 10529-10540 (2017).
26. Stalnaker, T. A., Cooch, N. K. & Schoenbaum, G. What the orbitofrontal cortex does not do. Nat. neurosci. 18, 620-627 (2015).
27. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal cortex as a cognitive map of task space. Neuron 81, 267-279 (2014).
28. Price, J. L. Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. Ann. N. Y. Acad. Sci. 1121, 54-71 (2007).

ACKNOWLEDGEMENTS
This work was supported by the European Union Seventh Framework Programme under grant agreement number 607310 (*Nudge-IT*), and the Netherlands Organisation for Health Research and Development (ZonMW) grant 912.14.093 (*Shining light on loss of control*).

AUTHOR CONTRIBUTIONS
J.P.H.V., R.A.H.A. and L.J.M.J.V. designed the experiments. J.P.H.V. performed the experiments and analyses. J.P.H.V., H.E.M.d.O., R.A.H.A. and L.J.M.J.V. wrote the manuscript.

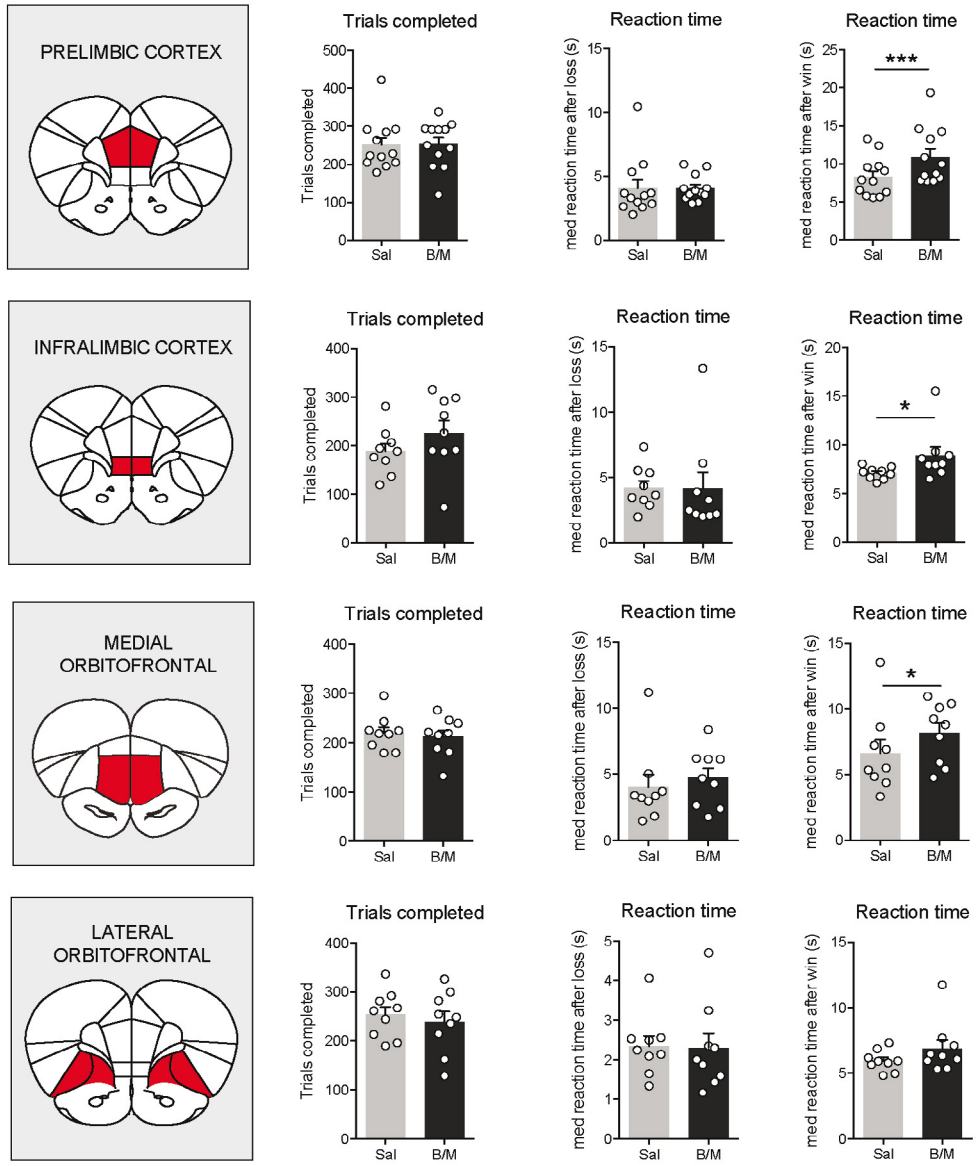
COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

SUPPLEMENTARY FIGURE 1



Acceptable infusion sites

SUPPLEMENTARY FIGURE 2



Additional measures of the task after pharmacological PFC inactivation. Reaction times after a win trial were generally higher than after a loss trial, since in win trials the new trial started immediately after the animals entered the food port (and spent time eating the pellet). * P < 0.05, *** P < 0.001 (post-hoc Holm-Sidak test).

SUPPLEMENTARY TABLE 1

	Model	Free parameters	Aggregate LL	$P_{explained}$	Aggregate AIC	Aggregate BIC	# of sessions best described by model	XP	PXP
1	Random	-	-12286.0	0.5000	24572.1	24572.1	0/78	0	0
2	Rescorla-Wagner 1	α, β	-10379.3	0.5568	21070.6	21596.3	19/78	0	0
3	Rescorla-Wagner 2	α+, α-, β	-10275.8	0.5600	21019.5	21808.1	7/78	0	0
4	Rescorla-Wagner 3	α+, α-, β, π	-9988.5	0.5692	20601.1	21652.4	46/78	1	1
5	Rescorla-Wagner-Pearce-Hall hybrid	α, β, π, η	-10065.7	0.5667	20755.3	21806.6	6/78	0	0

Abbreviations: LL, log-likelihood; Pexplained, fraction of choices explained by the model on every single trial (total trials on average ≈ 227); AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; XP, exceedance probability; PXP, protected exceedance probability.

CHAPTER 4

Differential contributions of striatal dopamine D1 and D2 receptors to value-based learning and decision making

Jeroen P.H. Verharen
Roger A.H. Adan*
Louk J.M.J. Vanderschuren*

* Equal contribution

Manuscript in preparation

Highlights

- We infused dopamine receptor (ant)agonists in different parts of the striatum during probabilistic reversal learning
- The ventral striatum is important for learning, and guides positive feedback learning through dopamine D1 and negative feedback learning through dopamine D2 receptors
- Exploration versus exploitation is mediated by dopamine D2 receptors in ventral and dorsolateral striatum

Techniques

Behavioral pharmacology

Computational modeling

CHAPTER 4

Dopamine is thought to have an important mediating role in value-based learning and decision making by signaling reward prediction errors and facilitating cognitive flexibility, motivation and movement. Dopamine receptors can roughly be divided into the D1 and D2 subtypes, and it has been hypothesized that in the striatum these two types of receptors have an antagonistic function in facilitating approach and avoidance behaviors, respectively. Here, we tested the contribution of these striatal dopamine receptors to the core processes underlying value-based learning and decision making in rats. By using computational trial-by-trial analysis of data of a probabilistic reversal learning task after systemic or local treatment with dopamine D1 and D2 receptor agonists and antagonists, we show that negative feedback learning is mediated by stimulation of the dopamine D2 receptor and positive feedback learning by stimulation of the dopamine D1 receptor in the ventral, but not dorsal, striatum. Furthermore, infusion of D2 agonist quinpirole in the ventral or dorsolateral, but not dorsomedial, striatum promoted explorative choice behavior, suggesting an additional function of these areas in value-based decision making. Together, these data support the idea that dopamine D1 and D2 receptors have a dissociable function in mediating positive and negative feedback learning, and provide evidence that dopamine facilitates value-based behaviors through distinct striatal regions.

1. Introduction

Many decisions we make in everyday life are the result of a process in which the expected gains and losses of different courses of action are weighed and compared, and these expectations are often based on the outcomes of similar actions taken in the past. The process by which these action-outcome associations are acquired and stored to guide future behavior, thereby linking positive and negative experiences to actions under different physiological states, is called reinforcement learning[1,2]. Deficits in this process have been implicated in a wide variety of mental conditions, including depression, mania, attention-deficit/hyperactivity disorder and addiction[3-9].

Dopamine (DA) is an important modulator of motivated behaviors, and it does so by attributing salience to relevant cues[10], by guiding movement[11], and by signaling reward prediction errors[12-14]. Especially this latter function of DA is thought to be fundamental for value-based learning. Reward prediction error theory posits that midbrain DA neurons signal a discrepancy between anticipated and received reward or punishment. As such, when a rewarding outcome is better than expected, phasic firing of DA neurons is transiently increased, while a worse-than-expected outcome triggers a reduction in firing. Downstream dopaminoreceptive brain areas can use these signals to update future expectations of actions, in order to adapt to environmental changes. Furthermore, DA has been implicated in cognitive flexibility, another core process involved in decision making, since manipulations of the DA system disrupt performance in tasks such as reversal learning and set shifting[15-18]. Importantly, many of the neuropsychiatric conditions that have been associated with deficits in learning, cognitive flexibility and decision making have also been linked to alterations in the DA system[19-25], and pharmacological treatment of many of these conditions is aimed at DAergic neurotransmission in the brain. Despite the wide use of these types of medication, little is known about how these drugs affect physiological processes in the brain, and

specifically if and how they change the computational mechanisms underlying decision making. Furthermore, it is has been suggested that the D1 subclass of striatal DA receptors is mainly involved in behavioral activation and appetitive learning, while the D2 subclass is important for behavioral inhibition and avoidance learning[26-28]. However, whether this is directly driven by antagonistic contributions to positive versus negative feedback learning remains unknown.

Here, we studied the role of DAergic neurotransmission in the striatum in value-based learning in rats using a probabilistic reversal learning paradigm[29]. By applying a computational Q leaning model[1,30-32] to the data, we tried to gain further insight into the strategy the animals used to perform the task, and thereby unravel how the two major subclasses of DA receptors contribute to the core processes underlying decision making. Specifically, we studied the effects of pharmacological activation and inactivation of the DA D1 and D2 receptors in the ventral striatum (VS), dorsolateral striatum (DLS) and dorsomedial striatum (DMS) on reward learning, punishment learning, choice perseveration, and choice stochasticity. We predicted an important role of DA receptors in the VS in reward and punishment learning, given its function in processing reward prediction error and facilitating motivation[18,33], and of DA receptors in the dorsal parts of the striatum in mediating cognitive flexibility, given its function in balancing goal-directed and habitual behaviors[34,35].

2. Materials and methods

2.1 Animals

A total of 68 adult male (>300 g) Long-Evans rats (Janvier labs, France) were used for the experiments. Rats were housed in pairs (for systemic drug treatment) or singly (for intracranial infusions) in a humidity- and temperature-controlled room and they were kept on a 12h/12h reversed day/night cycle (lights off at 8AM). All experiments took place in the dark phase of the animals. Animals were kept on food restriction (~4.5g standard lab chow per 100g body weight per day) during behavioral training and the experiments. All experiments were conducted in accordance with European (2010/63/EU) and Dutch (Wet op de Dierproeven, 2014) legislation, and approved by the Animal Ethics Committee and Animal Welfare Body of Utrecht University.

Six independent cohorts of animals were used for the experiments:
Cohort A (n = 15): systemic treatment with quinpirole, SCH23390 and SKF82958
Cohort B (n = 9): systemic treatment with raclopride, SCH23390 and SKF82958
Cohort C (n = 9): systemic treatment with quinpirole, raclopride and raclopride
Cohort D (n = 16): VS infusions
Cohort E (n = 6): DLS infusions
Cohort F (n = 13): DLS infusions (5) and DMS infusions (8)

Animals in cohort D, E and F were tested with all four drugs. In cohort D, the DA D2 receptor infusion experiment was conducted first, after which 5 animals were transferred to another experiment; the D1 infusion experiment was conducted in the remaining 11 animals.

2.2 Surgeries

The rats in cohorts D, E and F were equipped with guide cannulas aimed at the three investigated subregions of the striatum. Animals were anesthetized by an intramuscular injection of a cocktail of 10 mg/kg fluanisone and 0.3 mg/kg fentanyl (Hypnorm, Jansen Pharmaceutica, Belgium) and were subsequently placed in a stereotaxic apparatus. An incision was made along the midline of the skull and additional analgesia was applied by spraying lidocaine on the skull. Two small craniotomies were made above the area of interest, and two 26G guide cannulas (Plastics One, United States) were placed bilaterally above the region of interest. The used coordinates were:

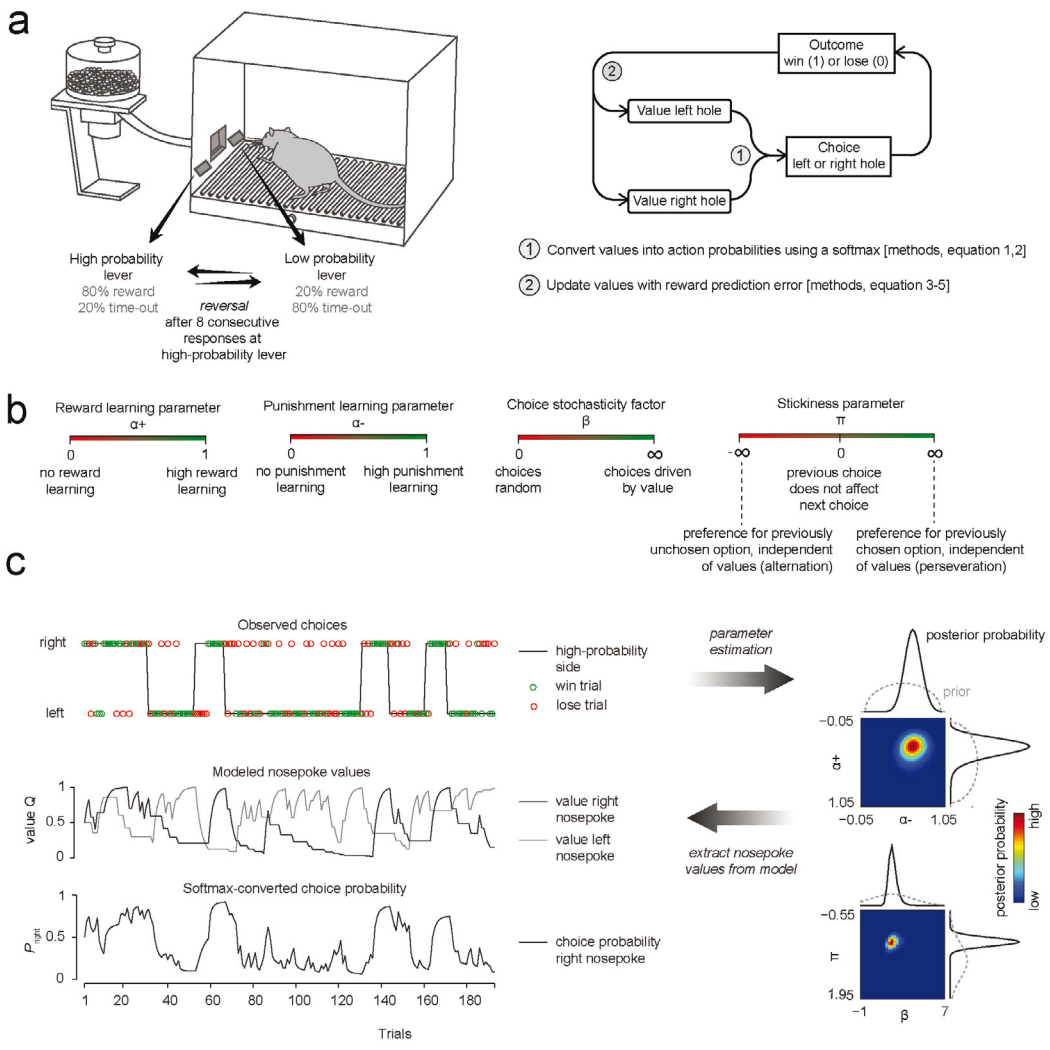


Figure 1 Task setup

a. Behavioral task and computational model.
b. Interpretation of computational model parameters
c. Trial-to-trial data of individual sessions was used to estimate the values of the parameters of the computational model. These model parameters describe the extent to which the trial outcomes affect the lever values (learning rates α^+ and α^-) and how these lever values are converted into action probabilities (stickiness parameter π and stochasticity parameter β).

VS		AP +1.2 mm	ML ±2.1 mm DV -6.3 mm from skull under a 5° angle
	or	AP +1.2 mm	ML ±2.7 mm DV -7.0 mm from skull under a 10° angle
DMS		AP +1.2 mm	ML ±2.3 mm DV -4.1 mm from skull under a 5° angle
DLS		AP +1.2 mm	ML ±3.4 mm DV -4.1 mm from skull

For the VS, 7 mm-long guide cannulas were used, and for the DLS and DMS 5 mm-long guide cannulas were used. Two different coordinates for the VS were used because we initially attempted to separately target the core and shell subregions of the nucleus accumbens, but we could not reliably determine whether the infused volume remained restricted to these subregions, especially since many of the infusions sites appeared on the border of the core and shell.

After lowering the guide cannulas to the desired coordinates, they were secured with screws, dental glue (C&B Metabond, Parkell, United States) and dental cement. The skin around the headcap was sutured and animals subsequently received carprofen for pain relief (5 mg/kg subcutaneously, for 3 days) and 5 ml saline for rehydration (subcutaneously, once). Dummy injectors were placed inside the guide cannulas and behavioral training started after a recovery period of 7 days.

2.3 Drugs, systemic injections and micro-infusions

The following drugs were used, which were all dissolved in sterile saline: (-)-quinpirole hydrochloride (Tocris Bioscience, United Kingdom), raclopride (Tocris Bioscience, United Kingdom), R(+)-SCH23390 hydrochloride (Sigma-Aldrich, The Netherlands), SKF82958 hydrobromide (Tocris Bioscience, United Kingdom).

For the systemic administration, all drugs were injected i.p., in a volume of 1 ml/kg, 15-20 minutes before the start of the behavioral task. Each drug was tested on three consecutive days, in a counterbalanced, within-subjects design. Doses were used that have been shown to elicit behavioral effects in other tasks in rats[36,37].

For the intracranial infusion experiments, the drugs (per side 1 µg of SCH23390, 5 µg of SKF82958, 7.5 µg of raclopride or 5 µg of quinpirole; based on previous experiments[38-40]) were infused in a volume of 0.5 µl per side, at a rate of 0.5 µl/min. Injectors protruded ~0.5 mm beyond the termination point of the guide cannulas during infusions. After infusion, the injectors were kept in place for an additional 30 seconds to allow diffusion of the solution into the tissue. Per DA receptor, infusion of the agonist, antagonist or saline was tested on three consecutive days, in a counterbalanced, within-subjects design.

2.4 Behavioral task

The behavioral task was conducted in operant chambers (Med Associates Inc., USA, 30.5×24.2×21.0 cm), which were placed in sound-attenuating cubicles. Operant chambers were equipped with a food port in which 45 mg sucrose pellets (SP; 5TUL, TestDiet, USA) could be delivered, flanked by two levers, and two cue lights above the levers. On the other side of the chamber there was a house light and a tone generator.

The behavioral task and training took place as described in ref. 32. In brief, animals could earn sucrose pellets by responding on two levers that each differed in the probability of being reinforced (Figure 1a). At task initiation, one lever was randomly assigned as the high-probability lever and pressing this lever had a 80% chance of being reinforced (a sucrose pellet delivery) and 20% chance of not being reinforced (a 10s time-out). The other lever was assigned as the low-probability lever, which gave 20% chance of being reinforced and 80% of not being reinforced. Initial assignment of the left and right lever as high- or low-probability was counterbalanced between animals. When the animal made 8 consecutive responses on the high-probability lever, a reversal in reward contingencies occurred, so that the previously high-probability lever became the low-probability lever and vice versa. The task terminated after 90 minutes.

MedPC software automatically registered, per trial, the choices of the animals (left

or right), the outcome of the trial (reinforced or not reinforced), the side of the high-probability lever (left or right), a timestamp of the start of the trial (time of lever protrusion) and a timestamp of the response (time of lever press). From these data, the following parameters were extracted using Matlab (R2014a, MathWorks Inc., United States): the number of trials completed in the 90-minute session; the median reaction time of the animals (computed by extracting the median value from all the reaction times, i.e., time of lever press minus time of trial start; note that the median was taken because reaction times were not normally distributed); the number of reversals per 100 trials (computed by dividing the total reversals by the number of trials completed, multiplied by 100); and the fraction of rewarded trials (i.e., the fraction of trials in which reward was obtained). The choice of the animals per trial (left or right) and the outcome of each trial (reinforced or not reinforced) were used to perform the computational analysis.

2.5 Computational model

We used computational modeling[1,30-32] to extract different subcomponents of reward-based decision-making from the raw behavioral data (Figure 1b), and used a reinforcement learning model which we have previously shown to be the best descriptor of behavior of rats in the task[32]. This model assumes that on every trial, the agent (in this case the rat) makes a choice based on a representation of the value of each of these levers. In most cases, the agent chooses the lever with the highest value Q on each trial t. The relationship between lever values Q_{left} and Q_{right}, and the probability that the agent chooses left or right ($p_{left,t}$, respectively $p_{right,t}$) lever in every trial is described by a softmax function:

$$p_{right,t} = \frac{\exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})}{\exp(\beta \cdot Q_{left,t} + \pi \cdot \phi_{left,t}) + \exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})} \quad (1)$$

and $p_{left,t} = 1 - p_{right,t}$ (2)

In this function, β is the softmax inverse temperature, which indicates how value-driven the agent's choices are. If β becomes very large, then the value function $\beta \cdot Q_{s,t}$ of the highest valued side becomes dominant, and the probability that the agent chooses that side approaches 1. Is β zero, then $p_{left,t} = p_{right,t} = e^0/(e^0 + e^0) = 0.5$ (π not taken into account), so that choice behavior becomes random. β is sometimes referred to as the explore/exploit parameter, where a low β favors exploration (i.e., sampling of all options) and a high β favors exploitation (i.e., choosing the option which has proven to be beneficial). Therefore, a decrease in β may reflect more explorative choice behavior, although a large decrease in β could also indicate a general disruption of behavior, i.e., that the animal chooses more randomly.

Factor π is a stickiness parameter that indicates a preference for the previously chosen ($\pi > 0$; perseveration) or previously unchosen ($\pi < 0$; alternation) option. Here, ϕ is a boolean with $\phi = 1$ if that hole was chosen in the previous trial, and $\phi = 0$ if not. For example, if the right lever was chosen in trial *t-1*, then $\phi_{right,t}$ becomes 1, and $\phi_{left,t}$ becomes 0. This adds a certain amount of the value of π to the value function of the lever in trial *t*, in addition to the lever's expected value $Q_{right,t}$.

For the first trial, both lever values were initiated at 0.5. After each trial, the value of the chosen lever was updated based on the trial's outcome according to a Q-learning rule:

$$Q_{s,t} = \begin{cases} Q_{s,t-1} + \alpha^+ \cdot \text{RPE}_t & \text{for win trials} \\ Q_{s,t-1} + \alpha^- \cdot \text{RPE}_t & \text{for lose trials} \end{cases} \quad (3)$$

with $\text{RPE}_t = outcome_t - Q_{s,t-1}$ (4)

so that

$$\text{RPE}_t = \begin{cases} 1 - Q_{s,t-1} & \text{for win trials} \\ 0 - Q_{s,t-1} & \text{for lose trials} \end{cases} \quad (5)$$

where $Q_{s,t-1}$ is the value of the chosen lever. Here, α^+ and α^- indicate the agent's ability to learn from positive (reinforcement; reward delivery), respectively negative (reward omission) feedback. The value of the unchosen side was not updated and thus retained its previous value.

2.6 Model fitting

The best-fit model parameters were estimated for each individual session, by calculating the probability of observing a certain choice sequence given a set of parameters {α^+, α^-, π, β}, by means of summing the logarithms of the probability of every observed choice from trial 1 to the last trial *T*:

$$\log(P(\text{data}\,|\,\text{model},\{\alpha^+,\alpha^-,\pi,\beta,\eta\})) = \sum_{t=1}^{T} \log(P(\text{choice}_t\,|\,Q_{left,t},Q_{right,t},\phi_{left,t},\phi_{right,t})) \quad (6)$$

We set weakly informative priors on the parameters to regularize the parameters towards realistic ones on a population level. The used priors were:

α^+, α^-	betapdf(1.5, 1.5)
π	normpdf(0.5, 0.5)
β	normpdf(2, 2)

Multiplication of these priors with the likelihood gives the posterior probability of the model parameters given the observed choice sequence:

$$P(\{\alpha^+,\alpha^-,\pi,\beta\}\,|\,\text{data,model}) = P(\text{data}\,|\,\text{model},\{\alpha^+,\alpha^-,\pi,\beta\}) \cdot P(\{\alpha^+,\alpha^-,\pi,\beta\}\,|\,\text{model}) \quad (7)$$

The posterior probability was calculated for many combinations of parameters {α^+, α^-, π, β}, and arranged in a 4-dimensional grid (Figure 1c). Best-fit parameter values were then estimated by integrating these posterior probabilities over the parameter's range, marginalized over the other three parameters (black lines next to the heatmaps in Figure 1c).

2.7 Histology

After the experiments, injection locations were histologically verified by an experimenter blind to the outcome of the behavioral experiments. Animals were first transcardially perfused by phosphate-buffered saline (PBS) followed by 4% paraformaldehyde in PBS. Brains were stored at 4°C, and were kept in 4% paraformaldehyde in PBS for at least 24 hours, followed by a 30% sucrose solution for at least 48 hours. Brains were sliced using a cryostat (50 µm) and were stained using a 5% Giemsa solution (Sigma-Aldrich, The Netherlands). One animal was excluded from the DMS group due to misplacement of the cannulas. See Figure 2 for an overview of the infusion sites.

2.8 Code accessibility

MedPC script of the probabilistic reversal learning task is available at github.com/jeroenphv/ReversalLearning.

2.9 Statistics

Statistical tests were performed with Prism 6 (GraphPad Software Inc.). For each systemically

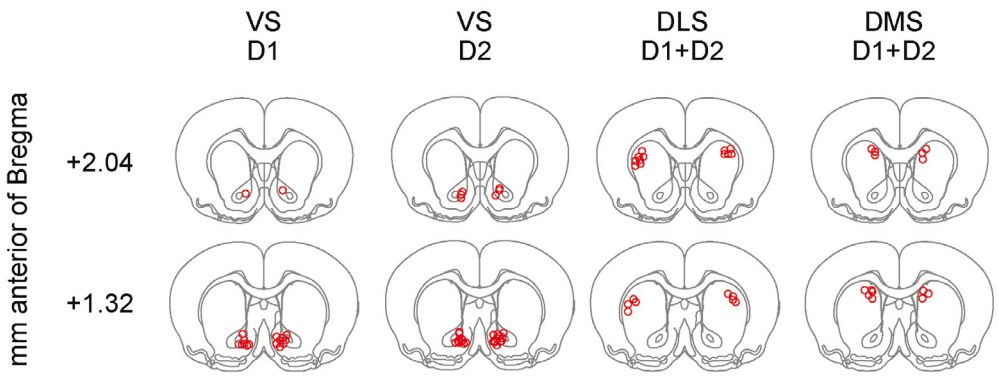


Figure 2 Infusion sites included in the analysis

tested drug, a 1-way repeated measures analysis of variance (ANOVA) with Greenhouse-Geisser correction was used to calculate significance. When the ANOVA yielded significant results ($p < 0.05$), a post-hoc LSD test was used to compare the high drug dose and the low drug dose with vehicle. For the intracranial infusion data, paired t-tests were performed in which the tested drugs were compared against vehicle. All statistics are presented in Supplementary Table 1. In all figures, the statistical range was denoted by the following symbols: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

3. Results

3.1 Systemic administration

Treatment with the DA D1 receptor antagonist SCH23390 significantly decreased the total number of trials completed (Figure 3a; for statistical measures, see Supplementary Table 1) and increased the reaction time. However, none of the parameters of the reinforcement learning model were significantly affected (Figure 3b), which was reflected by the lack of effect on the two general performance measures; total reversals per 100 trials and fraction of rewarded trials (Figure 3a).

Injection of the DA D1 receptor agonist SKF82958 reduced the number of completed trials and increased the response latency (Figure 3a). Additionally, it led to a numerically modest but significant decrease in reward learning rate α^+ (Figure 3b), but this had no consequences for the total reversals made by the animals or the fraction of rewarded trials (Figure 3a). No effects were observed on the value estimates of punishment learning parameter α^+, perseveration parameter π or choice stochasticity factor β (Figure 3b).

Treatment with the DA D2 receptor antagonist raclopride increased the response latency, without a significant effect on the total number of completed trials (Figure 3c). Furthermore, neither of the measures of task performance were affected (Figure 3c), which was reflected by the absence of effects on the computational model parameters (Figure 3d).

Injection of the DA D2 receptor agonist quinpirole decreased the number of completed trials and increased response latencies (Figure 3c). It also impaired task performance, both in terms of the total number of reversals and the fraction of rewarded trials (Figure 3c). Computational analysis revealed that this was associated with a combined decrease in reward learning rate α^+, which was numerically modest, and a decrease in punishment learning rate α^-, which was numerically larger than the effect on reward learning (Figure 3d). Moreover, choice stochasticity factor β was significantly reduced, but stickiness parameter π was not (Figure 3d).

3.2 VS infusions

Infusion of DA D1 receptor antagonist SCH23390 into the VS did not affect the total trials completed or the response latencies (Figure 4a). A significant increase in the total number of reversals was observed, but not in the fraction of rewarded trials (Figure 4a). However, none of the computational modeling parameters were significantly altered (Figure 4b), although a trend towards an increase in choice stochasticity parameter β was observed ($p = .07$; see also Supplementary Table 1).

Infusion of DA D1 receptor agonist SKF82958 did not significantly change the total trials completed, the response latency or the two measure of task performance (Figure 4a). However, a significant decrease was observed in the value estimate of reward learning parameter α^+, without effects on punishment learning rate α^-, stickiness parameter π or stochasticity factor β.

Infusion of DA D2 receptor antagonist raclopride into the VS significantly increased the animals' response latency, but did not change the total trials completed (Figure 4c), the two measures of task performance (Figure 4c) or any of the computational model parameters (Figure 4d).

In contrast, infusion of the DA D2 receptor agonist quinpirole affected different measures of task behavior. First, it strongly decreased the number of trials completed in the task and increased the response latency of the animals (Figure 4c). It also impaired task performance in terms of the total reversals achieved, but not in terms of the fraction of rewarded trials. This decreased number of reversals was driven by decreases in the value estimates of punishment learning parameter α^- and choice stochasticity factor β, but not by changes in reward learning parameter α^+ or stickiness parameter π (Figure 4d).

3.3 DLS infusions

Infusion of the DA D1 receptor antagonist SCH23390 or agonist SKF82958 into the DLS had no effect on any of the task measures (Figure 5a,b). In contrast, infusion of the DA D2 receptor antagonist raclopride significantly reduced the total number of completed trials and increased response latencies, but did not affect the two measures of task performance (Figure 5c). Consistently, none of the computational modeling parameters were significantly changed (Figure 5d). Infusion of the DA D2 receptor agonist quinpirole into the DLS increased the animals' response latency, but did not affect the total trials completed or any of the two performance measures (Figure 5c). It did, however, lead to a significant decrease in the value estimate of choice stochasticity factor β, without any effects on the other computational model parameters (Figure 5d).

3.4 DMS infusions

After infusion into the DMS, none of the drugs affected performance in the task or changed the value estimates of the computational model parameters (Figure 6a-d). Moreover, DA D1 antagonist SCH23390 and agonist SKF82958 did not affect the trials completed in the task or response latencies (Figure 6a). Infusion of the DA D2 receptor antagonist raclopride increased the response latency of the animals, but did not change the number of trials completed (Figure 6c). Conversely, infusion of the DA D2 receptor agonist quinpirole decreased the number of trials completed in the task without a significant effect on the animals' response latency (Figure 6c).

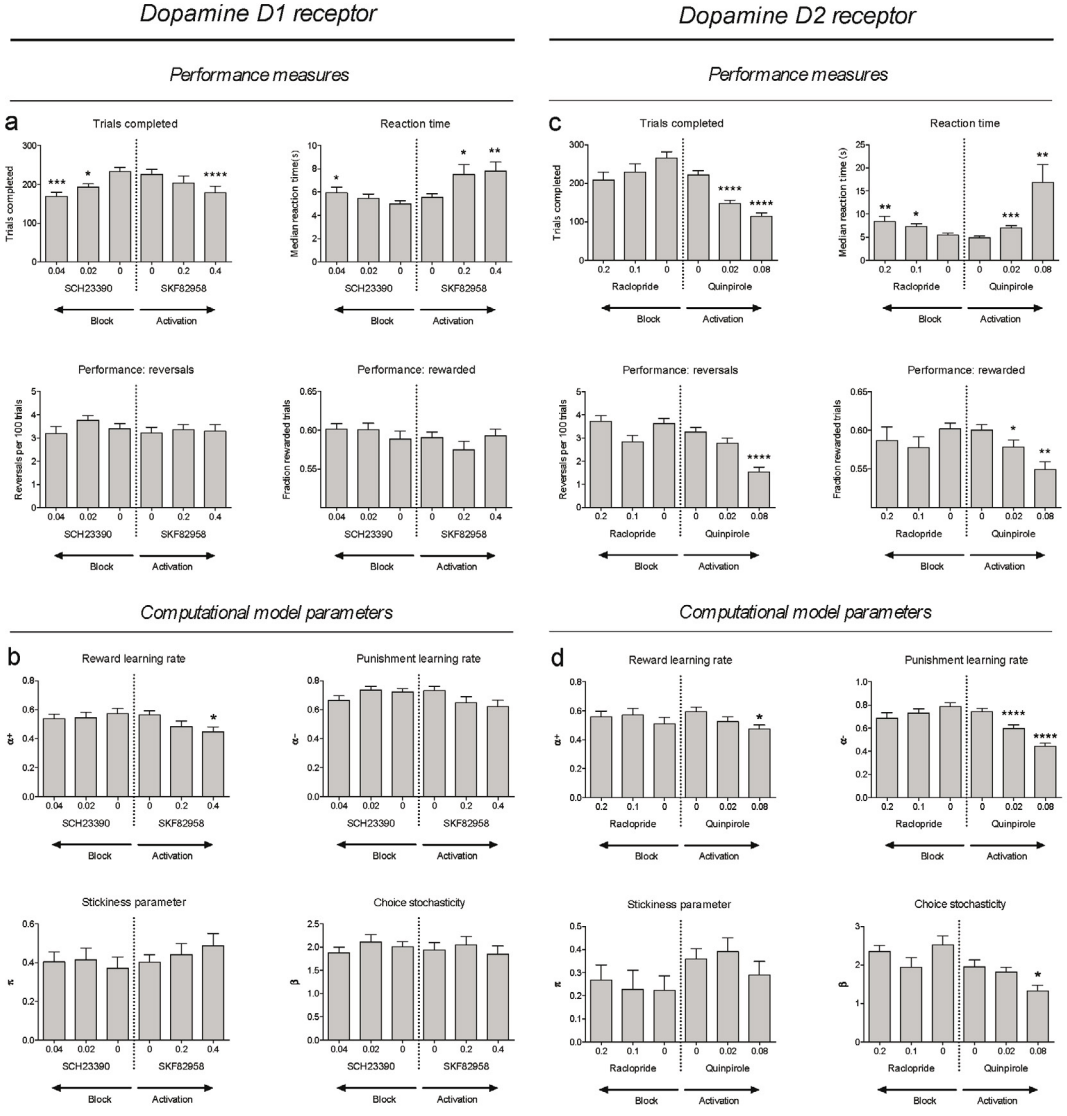


Figure 3 Systemic treatment with DA receptor (ant)agonists

a. Effects of systemic treatment with the DA D1 receptor antagonist SCH23390 (0, 0.02 or 0.04 mg/kg) and agonist SKF82958 (0, 0.2 or 0.4 mg/kg) on the behavioral measures of task performance. In all figures, the statistical range is denoted as: * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.

a. Effects of systemic treatment with the DA D1 receptor antagonist SCH23390 (0, 0.02 or 0.04 mg/kg) and agonist SKF82958 (0, 0.2 or 0.4 mg/kg) on the computational modeling parameters.

b. Effects of systemic treatment with the DA D2 receptor antagonist raclopride (0, 0.1 or 0.2 mg/kg) and agonist quinpirole (0, 0.02 or 0.08 mg/kg) on the behavioral measures of task performance.

c. Effects of systemic treatment with the DA D2 receptor antagonist raclopride (0, 0.1 or 0.2 mg/kg) and agonist quinpirole (0, 0.02 or 0.08 mg/kg) on the computational modeling parameters.

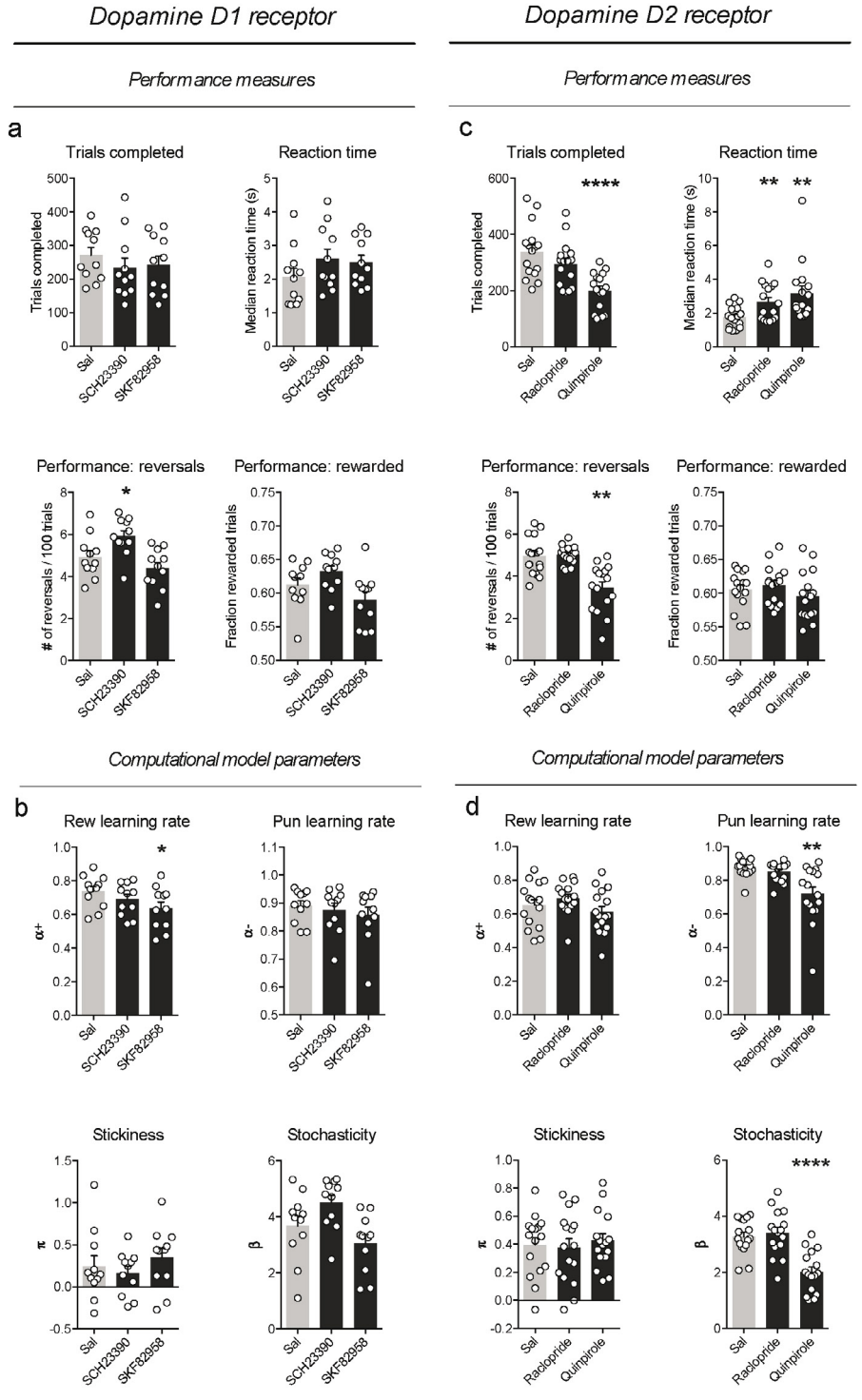


Figure 4 Ventral striatum infusions. **a,b.** Effects of intra-VS infusion of the DA D1 receptor antagonist SCH23390 (1 µg/side) and agonist SKF82958 (5 µg/side) on task performance. **c,b.** Effects of intra-VS infusion of the DA D2 receptor antagonist raclopride (7.5 µg/side) and agonist quinpirole (5 µg/side) on task performance.

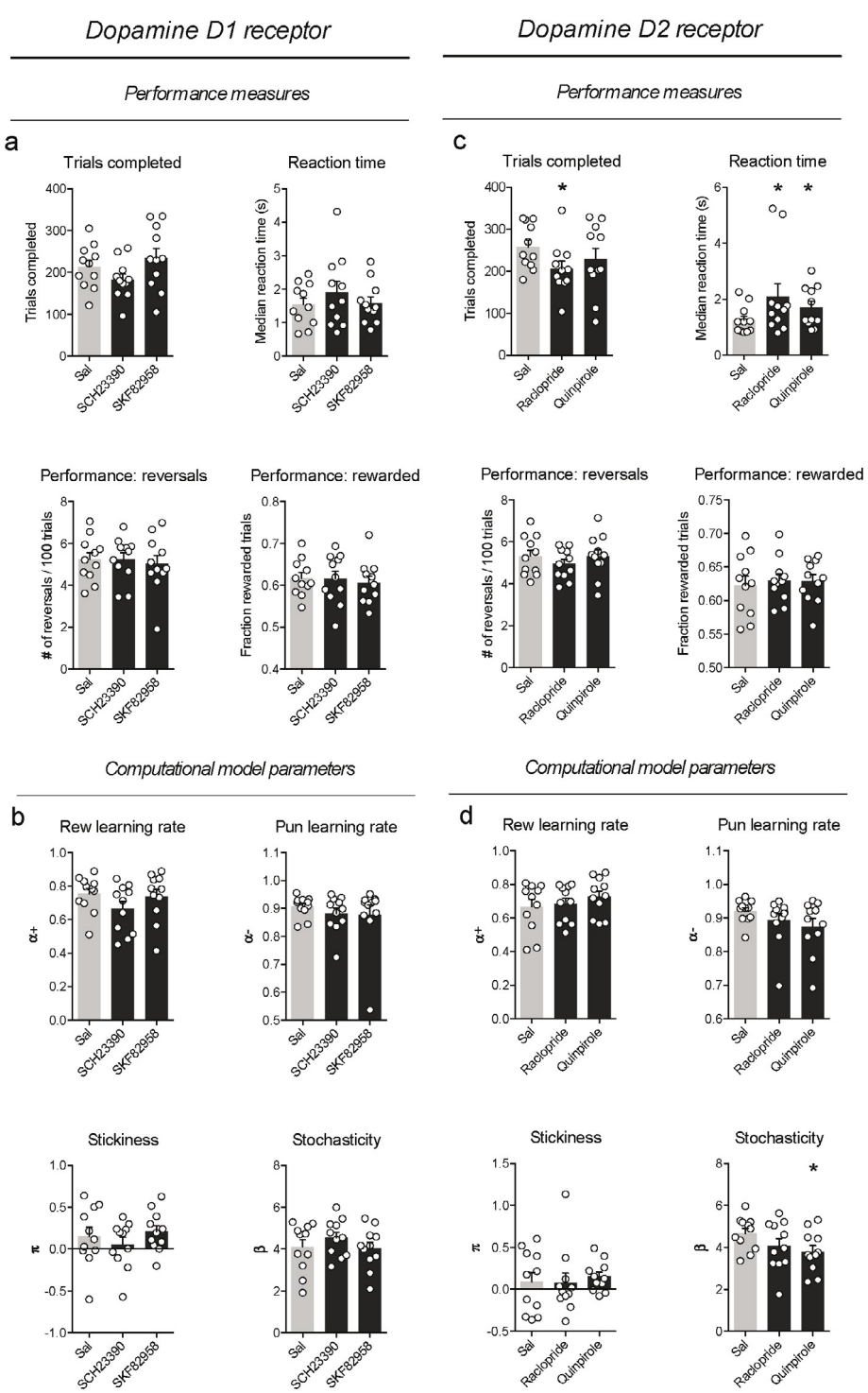


Figure 5 Dorsolateral striatum infusions. **a,b.** Effects of intra-DLS infusion of the DA D1 receptor antagonist SCH23390 (1 µg/side) and agonist SKF82958 (5 µg/side) on task performance. **c,b.** Effects of intra-DLS infusion of the DA D2 receptor antagonist raclopride (7.5 µg/side) and agonist quinpirole (5 µg/side) on task performance.

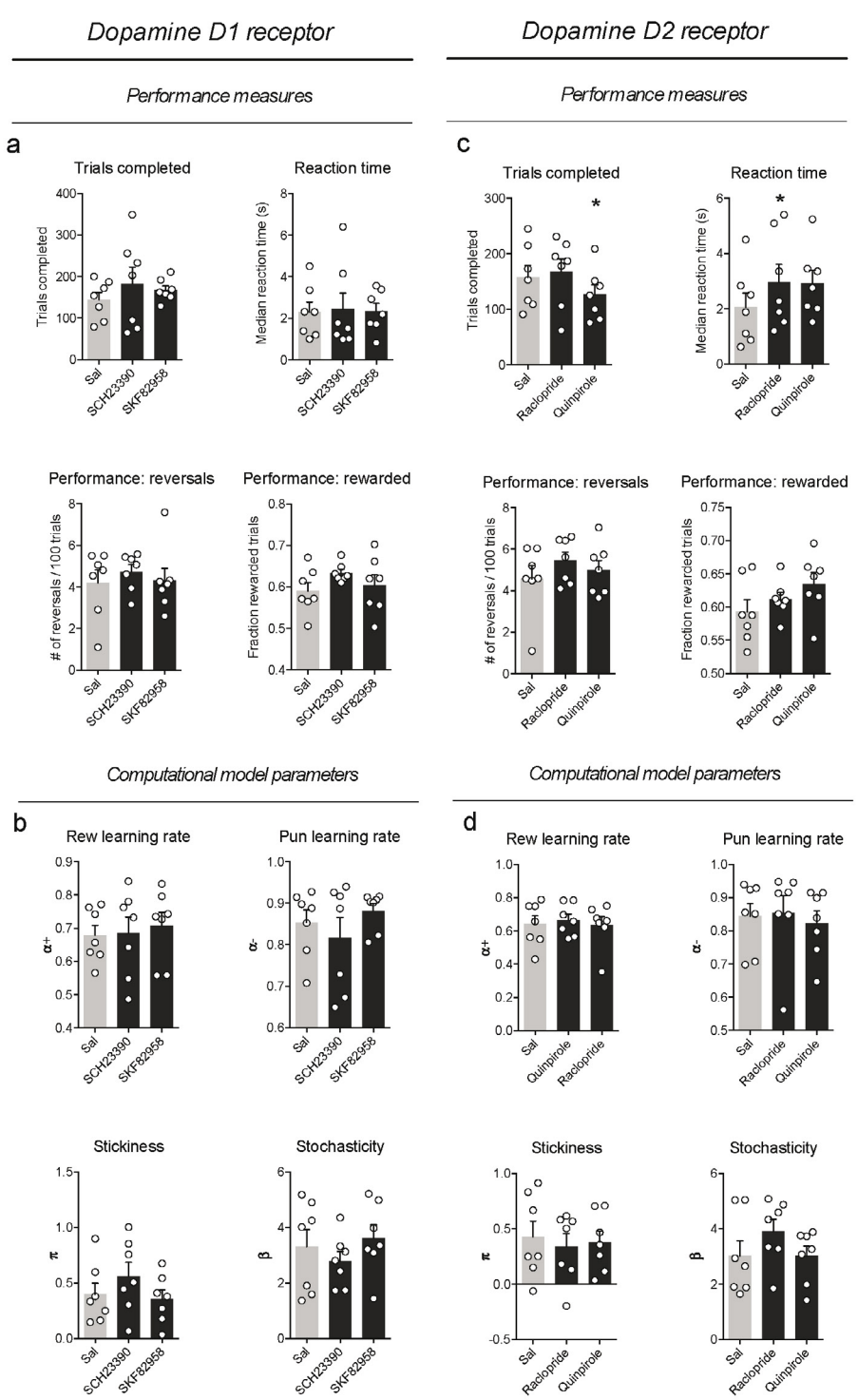


Figure 6 Dorsomedial striatum infusions. **a,b.** Effects of intra-DMS infusion of the DA D1 receptor antagonist SCH23390 (1 µg/side) and agonist SKF82958 (5 µg/side) on task performance. **c,b.** Effects of intra-DMS infusion of the DA D2 receptor antagonist raclopride (7.5 µg/side) and agonist quinpirole (5 µg/side) on task performance.

4. Discussion

4.1 Computational model

In this study, we have tested the effects of selective DA D1 and D2 receptor antagonists and agonists on serial probabilistic reversal learning, and used a computational reinforcement learning model to assess how these drugs impacted on subcomponents of value-based learning and decision making. We used a model that we have previously shown to be superior in explaining the rats' choice behavior during this task[32], which describes the behavior of the rats in terms of four parameters. The first two model parameters are reward and punishment learning rates α+ and α-, which indicate the extent to which a single reward delivery or omission impacts the value representation of the chosen lever. As such, a learning rate close to 1 indicates that a single trial outcome strongly affects the value of the chosen lever, while a learning rate close to 0 indicates that the value is barely updated on the basis of feedback (and the value representation is thus based on a longer history of outcomes). Stickiness parameter π indicates the amount of perseveration of the rats, and reflects the extent to which the animals have a preference for the lastly chosen lever; a positive value of π is indicative of perseveration on the same lever, a negative value of π is indicative of alternation between levers, whereas a π value close to 0 indicates that the lastly chosen lever does not affect the choice in the next trial (and hence choices are based on a comparison of the value estimates of the levers). Finally, stochasticity factor β reflects the extent to which the choices of the animals are driven by value; a high value of β indicates that the animals deterministically choose the highest-valued lever, while a β value close to 0 indicates sampling of both options (i.e., random choice). β is sometimes referred to as the explore/exploit parameter, where a high value of β favors exploitation of knowledge about the lever's values, whereas a low value of β favors exploration of both choice options.

This computational modeling approach informs about the strategy the animals use to perform the task, and demonstrates to what extent animals use rewarding and punishing feedback from the task to make choices. This analysis provides in-depth insights into the behavior of the animals besides the classical measures of performance, and sometimes reveals subtle changes in behavior that would not have been detected otherwise (e.g., the effects on reward learning in Figure 3b and 4b).

		Systemic	VS	DLS	DMS
DA D1 antagonist	SCH23390	↓↓ Trials completed ↑ Response latency			
DA D1 agonist	SKF82958	↓ Reward learning ↓↓ Trials completed ↑↑ Response latency	↓ Reward learning		
DA D2 antagonist	Raclopride	↑↑ Response latency	↑↑ Response latency	↓ Trials completed ↑ Response latency	
DA D2 agonist	Quinpirole	↓ Reward learning ↓↓ Punishment learning ↓ Stochasticity ↓↓ Trials completed ↑↑ Response latency	↓↓ Punishment learning ↓↓ Stochasticity ↓↓ Trials completed ↑↑ Response latency	↓ Stochasticity ↑ Response latency	↓ Trials completed ↑ Response latency

Table 1 Effects of DA (ant)agonists on the computational model (black) and motivational and motoric task parameters (gray).

4.2 Effects of systemic treatment with DA drugs

Findings of the systemic treatment experiment (Table 1) show a reduction in reward learning after systemic activation of the DA D1 and D2 receptors, but not after treatment with their respective antagonists. Punishment learning was solely dependent on the DA D2 receptor, as treatment with the agonist quinpirole (but not the D2 receptor antagonist raclopride) decreased the value estimate of this parameter. Furthermore, treatment with the DA D2 receptor agonist quinpirole decreased choice stochasticity factor β, indicating that animals shifted towards a decision-making strategy of exploration, rather than exploitation. No effects were observed on the value estimate of stickiness parameter π, suggesting that choice perseveration is not dependent on DA neurotransmission.

All drugs made the animals' responses after trial start significantly slower (i.e., an increased response latency). Furthermore, all drugs, except for DA D2 antagonist raclopride, decreased the total number of trials completed in the task, indicative of changed motivation, attention or hunger. The finding that the pattern of effects on response latency and trials completed in most cases matched a "U" or "inverted-U" shape (Figure 3a,c), suggests that DA receptors normally act at optimal levels, and that deviations from that optimum, either through blockade of these receptors with the antagonist or activation of these receptors with the agonist, impair behavior.

4.3 Striatal subregion-specific effects

The striatal infusion experiments suggested that the effect of systemic treatment with the DA D1 receptor agonist SKF82958 on reward learning was exerted in the VS, and not the dorsal parts of the striatum (Table 1). The effects of systemic quinpirole on the computational model parameters were only partially replicated in the micro-infusion experiments. First, the decrease in punishment learning after systemic injection of quinpirole was also observed after infusion into the VS, but not DLS and DMS. Second, the decreased value estimate of choice stochasticity factor β was also seen after infusion of quinpirole in the VS and DLS, but not the DMS. Finally, the effect of systemic quinpirole treatment on reward learning was not observed after infusion of this agonist into either of the three striatal regions, suggesting that these effects were driven by DA D2 receptor stimulation elsewhere in the brain, for example in the prefrontal cortex or through D2 autoreceptors on midbrain DA neurons. Stimulation of these latter receptors would inhibit activity of DA neurons, thereby decreasing DA release, and thus preventing a peak in DA release during positive reward prediction[16,41], which may logically explain the observed decrease in positive feedback learning after systemic quinpirole injection. Furthermore, the effects of systemic treatment with DA D2 receptor-acting drugs on the response latency and trials completed were also seen after infusion of these drugs into the different parts of the striatum. However, this was not the case for the effects seen after systemic treatment with DA D1 receptor-acting drugs, suggesting that the effects of these drugs on these motivational and motoric task parameters were the result of the combined effects of these drugs in the striatal subregions, or that the effects arose from other dopaminoreceptive brain areas, like the prefrontal cortex or amygdala.

The differential effects of the DA D1 and D2 receptor agonists in the VS on reward versus punishment learning suggest that these forms of learning are segregated in the VS DA system, in that they are mediated by different cell types. Punishment learning, theoretically guided by negative prediction errors coded by DA neurons, seems only dependent on the VS DA D2 receptor, while reward learning, as guided by positive DAergic prediction errors, is dependent on the VS DA D1 receptor. This finding is in line with theoretical neurocomputational models of the basal ganglia that implicates striatal DA D1 receptor-expressing neurons (through the "direct Go-pathway") in reward sensitivity, and striatal D2 receptor-expressing neurons (through the "indirect NoGo-pathway") in punishment sensitivity[26-28,41,42]. Our lab recently showed that an abundance of DA in the brain, induced by systemic treatment with dopaminomimetics or by chemogenetic activation of VTA neurons projecting to the VS,

evokes a mental state that is characterized by insensitivity to loss and punishment[18]. Here, we provide evidence that this phenomenon is driven by overstimulation of VS D2 receptors and the subsequent impairment in adapting to negative feedback. Indeed, the learning effects we observed after systemic injection or intra-VS infusion of D1 and D2 agonists were numerically larger for the D2 agonist than for the D1 agonist, which may explain why overstimulation of the VS with DA itself affects punishment learning, rather than reward learning.

The systemic effects of D2 receptor agonist quinpirole on choice stochasticity factor β were driven by action of this drug in the VS and DLS, but not the DMS. β is sometimes referred to as the explore/exploit parameter, and a decrease in its value indicates that the choices of the animals were more explorative by nature than under baseline conditions, thus being less driven by the value of the two levers. As such, the amount of exploration versus exploitation is a descriptor of behavior that is related to value-based decision making, rather than value-based learning. Although relatively little is known about the neural basis of this aspect of decision making[43], it has been shown that in humans, fMRI bold responses in the striatum (as well as in the ventromedial prefrontal cortex) are related to exploitative decisions (i.e., choosing the highest valued option)[44]. Furthermore, it has recently been shown that the balance between exploration and exploitation in human subjects is related to two genes linked to DAergic function[45]. That said, although the observed increase in the amount of exploratory choices is indicative of changes in value-based decision making, it must be noted that a decrease in β could also reflect a general disruption of behavior, thereby inducing more random choice behavior, for example because of a memory deficit or attentional impairment. However, this is not very likely given the absence of effects of quinpirole treatment on the stickiness parameter, as well as the absence of effects on reward learning after VS infusion of quinpirole and the absence of any changes in learning after DLS infusion of quinpirole.

It is interesting to note that treatment with DA receptor agonists affected learning rates by reducing these (i.e., impair learning). This indicates that DA operates on an optimal level, and that upward deviations from this optimum can evoke learning impairments. This is somewhat conflicting with the aforementioned model that states that D1 receptor activation would enhance reward sensitivity through activation of the "direct Go-pathway" in the striatum and D2 receptor blockade would enhance punishment sensitivity through the "indirect NoGo-pathway"[42]. One possibility is that the agonist occupies the DA receptors, thereby not allowing D1 or D2 receptor-expressing cells to detect transient changes in DA release during reward prediction errors.

Considering the role of DA in reward-based learning, it is surprising that no changes in performance or learning rates were observed after treatment with either the DA D1 receptor antagonist SCH23390 or the D2 receptor antagonist raclopride, even though we used concentrations with which we have observed behavioral effects in the past[36,37]. It could reasonably be argued that higher concentrations than the ones we have used may distort learning, but the animals became disengaged in the task after treatment with concentrations higher than 0.04 mg/kg SCH23390 or 0.2 mg/kg raclopride, thereby not completing enough trials to draw reliable conclusions about task performance (data not shown). Interestingly, changes in learning and decision-making have been observed after local micro-infusions of DA D1 and D2 antagonist into the brain, suggesting that blockade of DA receptors, both in the striatum[46] and the prefrontal cortex[17,47], has the potential to disrupt behavior in certain tasks. Thus, systemic treatment with DA antagonists might have affected motivation or motor behavior *before* it affected task performance, thereby not allowing the detection of a learning deficit. That said, given that micro-infusion of the DA D1 and D2 receptor antagonists in the striatum also did not affect the computational model parameters suggests that activity of an isolated class of DA receptors is not essential for reversal learning, for example because its function can partially be taken over by the other class of receptors. Importantly, the aforementioned effects of DA antagonists in the striatum on decision making[46] were observed in a task that entailed risky choice, and may rely on other cognitive processes than the task used in our study.

4.4 Concluding remarks

Value-based decision making is a fundamental process for an organism to thrive and survive in a changeable environment, and DA has been widely implicated in this process. Although claims have been made about the relative contributions of different brain regions and subclasses of DA receptors to reinforcement learning, it has never been systematically studied and dissected into its core processes. Here, we used a pharmacological and computational approach to investigate how the two main subclasses of DA receptors, the D1 and D2 receptors, contribute to four important components of value-based decision making: reward learning, punishment learning, choice perseveration, and choice stochasticity. Our research confirms previous notions of the involvement of the DA D2 receptor in avoidance behavior and the D1 receptor in approach behavior, and show that this may be driven by mediating fundamental value-based learning processes. Moreover, we show that treatment with DA receptor agonists more strongly affects behavior in the task than antagonists do, suggesting that activation of DA receptors disrupts the computational mechanisms driving reversal learning. Further experimental investigations will be needed to decipher the relative contributions of these receptors to value-based decision making in other brain regions, including the prefrontal cortex and amygdala, and study which downstream circuits process the learning signals that eventually establish complex choice behavior.

Acknowledgements

This work was supported by the European Union Seventh Framework Programme under grant agreement number 607310 (*Nudge-IT*)

Conflict of interest

None of the authors have financial interests or potential conflict of interest to report.

References

1. Sutton, R. S. & Barto, A. G. Reinforcement learning: An introduction. (MIT press, 1998).
2. Dayan, P. & Daw, N. D. Decision theory, reinforcement learning, and the brain. Cogn. Affect Behav. Neurosci. 89 (2008).
3. Roger, R. D. et al. Dissociable Deficits in the Decision-Making Cognition of Chronic Amphetamine Abusers, Opiate Abusers, Patients with Focal Damage to Prefrontal Cortex, and Tryptophan-Depleted Normal Volunteers: Evidence for Monoaminergic Mechanisms. Neuropsychopharmacology 20, 322-339 (1999).
4. Grant, S., Contoreggi, C. & London, E. D. Drug abusers show impaired performance in a laboratory test of decision making. Neuropsychologia 38, 1180-1187 (2000).
5. Murphy, F. C. et al. Decision-making cognition in mania and depression. Psychol. Med. 31, 679-693 (2001).
6. Ernst, M. & Paulus, M. P. Neurobiology of decision making: a selective review from a neurocognitive and clinical perspective. Biol. psych. 58, 597-604 (2005).
7. Johnson, S. L. Mania and dysregulation in goal pursuit: a review. Clin. Psychol. Rev. 25, 241-262 (2005).
8. Garon, N., Moore, C. & Waschbusch, D. A. Decision making in children with ADHD only, ADHD-anxious/depressed, and control children using a child version of the Iowa Gambling Task. J.of Att. Dis. 9, 607-619 (2006).
9. Noel, X., Brevers, D. & Bechara, A. A neurocognitive approach to understanding the neurobiology of addiction. Curr. opin. in neurobiol. 23, 632-638 (2013).

10. Berridge, K. C. The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology 191, 391-431 (2007).
11. Alexander, G. E. & Crutcher, M. D. Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends in Neurosci. 13, 266-270 (1990).
12. Schultz, W., Dayan, P. & Montague, P. R. A Neural Substrate of Prediction and Reward. Science 275, 1593-1601 (1997).
13. Keiflin, R. & Janak, P. H. Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. Neuron 88, 247-263 (2015).
14. Schultz, W. Dopamine reward prediction-error signalling: a two-component response. Nat. rev. Neurosci. 17, 183-195 (2016).
15. Clatworthy, P. L. et al. Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory. J. Neurosci. 29, 4690-4696 (2009).
16. Cools, R. et al. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J. Neurosci. 29, 1538-1543 (2009).
17. Floresco, S. B. Prefrontal dopamine and behavioral flexibility: shifting from an "inverted-U" toward a family of functions. Front Neurosci. 7, 62 (2013).
18. Verharen, J. P. H. et al. A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. Nat. comm. 9 (2018).
19. Nestler, E. J. & Carlezon, W. A., Jr. The mesolimbic dopamine reward circuit in depression. Biol. Psych. 59, 1151-1159 (2006).
20. Berk, M. et al. Dopamine dysregulation syndrome: implications for a dopamine hypothesis of bipolar disorder. Acta Psychiatrica Scandinavica 116, 41-49 (2007).
21. Cousins, D. A., Butts, K. & Young, A. H. The role of dopamine in bipolar disorder. Bipolar disorders 11, 787-806 (2009).
22. Russo, S. J. & Nestler, E. J. The brain reward circuitry in mood disorders. Nat. rev. Neurosci. 14, 609-625 (2013).
23. Nutt, D. J., Lingford-Hughes, A., Erritzoe, D. & Stokes, P. R. The dopamine theory of addiction: 40 years of highs and lows. Nat. rev. Neurosci. 16, 305-312 (2015).
24. Volkow, N. D. & Morales, M. The Brain on Drugs: From Reward to Addiction. Cell 162, 712-725 (2015).
25. Han, M. H. & Nestler, E. J. Neural Substrates of Depression and Resilience. Neurotherapeutics 14, 677-686 (2017).
26. Surmeier, D. J., Ding, J., Day, M., Wang, Z. & Shen, W. D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. Trends Neurosci. 30, 228-235 (2007).
27. Collins, A. G. E. & Frank, M. J. Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol. rev. 121, 337 (2014).
28. Francis, T. C. & Lobo, M. K. Emerging Role for Nucleus Accumbens Medium Spiny Neuron Subtypes in Depression. Biol. Psych. 81, 645-653 (2017).
29. Bari, A. et al. Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. Neuropsychopharmacology 35, 1290-1301 (2010).
30. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 2, 64-99 (1972).
31. Gershman, S. J. Empirical priors for reinforcement learning models. Journal of Math. Psychol. 71, 1-6 (2016).
32. Verharen, J. P. H., Kentrop, J., Vanderschuren, L. J. M. J. & Adan, R. A. H. Reinforcement learning across the rat estrous cycle. Psychoneuroendocrinology 100: 27-31 (2019).
33. Floresco, S. B. The nucleus accumbens: an interface between cognition, emotion, and action. Annu. Rev. Psychol. 66, 25-52 (2015).
34. Voorn, P., Vanderschuren, L. J., Groenewegen, H. J., Robbins, T. W. & Pennartz, C. M. Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci. 27, 468-474 (2004).
35. Balleine, B. W., Delgado, M. R. & Hikosaka, O. The role of the dorsal striatum in reward and decision-making. J. Neurosci. 27, 8161-8165 (2007).
36. van Gaalen, M. M., van Koten, R., Schoffelmeer, A. N. & Vanderschuren, L. J. Critical involvement of dopaminergic neurotransmission in impulsive decision making. Biol. Psych. 60, 66-73 (2006a).
37. van Gaalen, M. M., Brueggeman, R. J., Bronius, P. F., Schoffelmeer, A. N. & Vanderschuren, L. J. Behavioral disinhibition requires dopamine receptor activation. Psychopharmacology 187, 73-85 (2006b).
38. Wan, F. & Swerdlow, N. Intra-accumbens infusion of quinpirole impairs sensorimotor gating of acoustic startle in rats. Psychopharmacology 113, 103-109 (1993).
39. Cools, A. R., Miwa, Y. & Koshikawa, N. Role of dopamine D1 and D2 receptors in the nucleus accumbens in jaw movements of rats: a critical role of the shell. Eur. J. of Pharmacol. 286: 41-47 (1995).
40. Pattij, T., Janssen, M. C. W., Vanderschuren, L. J. M. J., Schoffelmeer, A. N. M. & Van Gaalen, M. M. Involvement of dopamine D 1 and D 2 receptors in the nucleus accumbens core and shell in inhibitory response control. Psychopharmacology 191, 587-598 (2007).
41. Cools, R. Role of dopamine in the motivational and cognitive control of behavior. Neuroscientist 14, 381-395 (2008).
42. Frank, M. J. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J. of Cogn. Neurosci. 17, 51-72 (2005).
43. Cohen, J. D., McClure, S. M. & Yu, A. J. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philos. Transactions of the Royal Soc. B: Biol. Sci. 362, 933-942 (2007).
44. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. Nature 441, 876-879 (2006).
45. Gershman, S. J. & Greshake Tzovaras, B. Dopaminergic genes are associated with both directed and random exploration. bioRxiv, doi:10.1101/357251 (2018).
46. Stopper, C. M., Khayambashi, S. & Floresco, S. B. Receptor-specific modulation of risk-based decision making by nucleus accumbens dopamine. Neuropsychopharmacology 38, 715-728 (2013).
47. St Onge, J. R., Abhari, H. & Floresco, S. B. Dissociable contributions by prefrontal D1 and D2 receptors to risk-based decision making. J. Neurosci. 31, 8625-8633 (2011).

CHAPTER 5

Reinforcement learning across the rat estrous cycle

Jeroen P.H. Verharen
Jiska Kentrop
Louk J.M.J. Vanderschuren*
Roger A.H. Adan*

* Equal contribution

Published in Psychoneuroendocrinology 100: 27-31 (2019)

Highlights

- We used computational modeling of data of female rats in a reward learning task
- Effects of estrous cycle on the core processes of reinforcement learning were studied
- Reward learning, exploration and motivation fluctuate across the cycle

Techniques

Computational modeling

CHAPTER 5

Reinforcement learning, the process by which an organism flexibly adapts behavior in response to reward and punishment, is vital for the proper execution of everyday behaviors, and its dysfunction has been implicated in a wide variety of mental disorders. Here, we use computational trial-by-trial analysis of data of female rats performing a probabilistic reward learning task and demonstrate that core computational processes underlying value-based decision making fluctuate across the estrous cycle, providing a neuroendocrine substrate by which gonadal hormones may influence adaptive behavior.

Introduction

Reinforcement learning is an essential mechanism for organisms to adapt to a dynamic environment, by allowing flexible alterations in behavior in response to positive and negative feedback, for example during foraging and social encounters[1]. As such, deficits in reinforcement learning have been implicated in several psychiatric conditions, including addiction and schizophrenia[2]. Given the large gender differences in the prevalence of mental disorders, and the existence of cyclic changes in the severity of schizophrenia and sensitivity to drugs in women[3], we sought to determine how the estrous cycle of females affects the computational processes that underlie reinforcement learning. To this aim, we tested a cohort of female rats on a probabilistic reversal learning paradigm[4,5], used computational modeling to extract the subcomponents of value-based decision making, and assessed how these components were affected by the estrous cycle.

Methods

Animals

Female, nulliparous Long-Evans rats (bred in-house; background Rj:Orl, Janvier labs, France; *n* = 30) weighing 180-220 gram were used for the experiment. Animals were tested for 10 consecutive days, to ensure that we had at least one measurement of every cycle stage per animal. Eventually, 5 animals had to be excluded because the cycle could not reliably be estimated or not all stages of the cycle were captured due to unreliable vaginal smears, leaving a final group of *n* = 25. Animals were socially housed in groups of 2-4 and kept on a reversed day/night cycle (lights on at 8AM), and behavioral experiments took place between 9AM and 1PM. During the training phase of the experiment, animals were kept on a food restriction regimen of 5 gram chow per 100 gram body weight, and during the 10 experimental days the animals were food restricted for 16 hours prior to the behavioral task. For the male group of animals (*n* = 18), that is included for comparison, Long-Evans rats (bred in-house; background Rj:Orl, Janvier labs, France) of roughly the same age, weighing 310-390 gram, were used. Animals had *ad libitum* access to water, except during behavioral experiments. The experiments were carried out in accordance with European Union guidelines (2010/63/EU), and approved by the Animal Welfare Body of Utrecht University and the Dutch Central Animal Testing Committee.

Behavioral task

The probabilistic reversal learning task (Figure 1a) took place in operant conditioning chambers (Med Associates Inc., USA) equipped with a food receptacle (with infra-red entry detection) flanked by two retractable levers and two cue lights, a house light and an auditory tone generator. One lever was randomly assigned as the high-probability lever, responding on which was reinforced (i.e. delivery of a sucrose pellet) with an 80% probability and not reinforced (i.e. a time-out) with a 20% probability. The other lever was assigned as the low-probability lever, responding on which had a 20% chance of being reinforced. Every single response on the high-probability and low-probability lever was reinforced with a 80% or 20% probability, respectively, irrespective of the outcome of the previous trials.

The session lasted for 60 minutes, and animals were constrained in the number of trials they could make only by the length of the session (maximum ~600 trials per session possible). A trial commenced by the illumination of the house light, and the presentation of the two levers into the operant cage. After a lever press by the animal, the levers retracted and the house light was turned off. For reinforced trials, a 45mg sucrose pellet (5TUL, TestDiet, USA) was delivered into the food port, and both cue lights that flanked the food receptacle were illuminated, and an auditory tone was played for 0.5s. A new trial commenced directly when the animal entered the food port (detected by the infra-red movement detector); this was signaled to the animal by extinction of the cue lights, illumination of the house light and presentation of the two levers. On non-reinforced trials, no additional cues were presented, leaving the animals in the dark during a 10s period.

Every time the animal made 8 consecutive responses on the high-probability lever, a reversal in reinforcement contingencies occurred, so that the high-probability and low-probability levers switched. This reversal was not signaled to the animal, so it had to infer this contingency switch from the outcomes of the trials.

The software automatically registered the responses and response times of the animals, as well as the outcome of the trial (reinforced or not), and the position of the high-probability lever.

Training

Animals first received lever press training, during which both levers were continuously presented, and a lever press was reinforced under a fixed ratio-1 schedule. When all animals made more than 50 lever presses in a session, the group progressed to the next phase of lever press training, in which randomly the left lever, the right lever, or both levers were presented to the animals, and pressing either lever was reinforced under a fixed ratio-1 schedule. In this phase of training, levers retracted after a response, and animals were subjected to the same sequence of events as during a reinforced trial in the probabilistic reversal learning task. When all animals made at least 100 responses in a session during this phase, the group received 6 training sessions of the probabilistic reversal learning task, before the experimental phase began (both females and males received these 6 training sessions in the final stage).

Estrous cycle determination

To determine the circulating levels of female sex hormones throughout the estrous cycle, vaginal smears were obtained for all test days between 11AM and 1PM, 1-2h after each test. Vaginal smears were collected by inserting the head of a sterile plastic smear loop (1µL; VWR, USA) and gently swabbing the vaginal wall. The collected cells were transferred to a drop of water on a glass microscope slide, air-dried and stained with 5% Giemsa (Sigma-Aldrich, The Netherlands) dissolved in water. Microscopic evaluation of the cells present in the vaginal smears was used to determine the phase of the estrous cycle[6,7] (Figure 1b). This was performed by a trained observer who was blind to smears from previous days and the behavioral data, and the following four parameters were estimated: the relative amount of cells present (on a scale from 1 to 5), and the percentage of nucleated cells, anucleated cells and leukocytes. Based on these four parameters and taking into account all 10 days, smears were assigned as proestrus, estrus or metestrus-diestrus, according to references 6 and 7. In brief; smears containing predominantly nucleated cells were assigned as proestrus, smears containing predominantly anucleated cells were assigned as estrus and smears containing leukocytes were assigned as metestrus-diestrus. Smears containing a combination of cells indicating a transition between phases were interpreted based on smears from neighboring days and references 6 and 7. Females that did not show a regular cycle over the course of 10 days were excluded from the analysis. If a single smear was unreliable for a given

day, but smears of neighboring days showed a predictable pattern coherent with a regular estrous cycle, the phase of the missing day was estimated; if not, that particular day was not included in the analysis.

Reinforcement learning model

The trial-by-trial data of every individual session was fit to a reinforcement learning model, which was a modification of the classic Rescorla-Wagner model[8], which assumes that the animals dynamically track the value of the outcome of responding on each of the two levers by incorporating positive (reward delivery) and negative (reward omission) feedback (Figure 1c,d). When learning from feedback is high ($\alpha \rightarrow 1$), these lever values are strongly dependent on the outcome of the last trial, but when learning is low ($\alpha \rightarrow 0$), lever values are based on an extended history of trials (thus the impact of a single reward delivery or reward omission on lever value is small). The model further incorporates the animals' preference for the lastly chosen lever, independent of lever values, which is captured by perseveration parameter π. Moreover, it incorporates stochastic choice, to distinguish between deterministic choice of the highest valued lever ($\beta \rightarrow \infty$) and a more exploratory sampling approach ($\beta \rightarrow 0$). Random effects model selection indicated that this modified Rescorla-Wagner model was able to predict the highest amount of observed choices compared to a set of other reinforcement learning models that we tested, including the classic Rescorla-Wagner model[8], a Pearce-Hall-Rescorla-Wagner hybrid model[9], and a win-stay, lose-switch model[10] (Supplementary Table 1).

The expected reward values of both levers, Q_{left} and Q_{right}, ranged from 0 (pressing the lever is never reinforced) to 1 (pressing the lever is always reinforced). Both lever values were initiated at a value of 0.5, and the value of the chosen lever Q_{chosen} was updated after every trial t based on the outcome of that trial:

$$Q_{chosen,t} = \begin{cases} Q_{chosen,t-1} + \alpha^{+} \cdot \delta_{t-1} & \text{for rewarded trials} \\ Q_{chosen,t-1} + \alpha^{-} \cdot \delta_{t-1} & \text{for time-out trials} \end{cases}$$

Here, α+ is the reward learning rate (learning from positive feedback), and α- is the punishment learning rate (learning from negative feedback), which range from 0 (no learning) to 1 (lever value completely determined by last outcome). δ_{t-1} represents the reward prediction error after the last trial *t-1*, so that:

$$\delta_{t-1} = \begin{cases} 1 - Q_{chosen,t-1} & \text{for rewarded trials} \\ 0 - Q_{chosen,t-1} & \text{for time-out trials} \end{cases}$$

Note that reward prediction error δ is negative for non-reinforced trials (outcome is lower than expected) and positive for reinforced trials (outcome is higher than expected). The value of the unchosen lever was not updated. Separate learning rates were used for learning from positive feedback (i.e. $\delta > 0$; rewarded trials) versus negative feedback (i.e. $\delta < 0$; time-out trials), so that changes in reward or punishment learning could be discerned in isolation

At the start of each trial, lever values Q_{left} and Q_{right} were converted to action probabilities using a Softmax function, so that the probability of choosing the right lever $p_{right,t}$ at trial t was given by the function:

$$p_{right,t} = \frac{\exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})}{\exp(\beta \cdot Q_{left,t} + \pi \cdot \phi_{left,t}) + \exp(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t})}$$

Here, β is the inverse temperature of the Softmax function, which is a measure for the extent to which the animal consistently chooses the highest valued lever ($\beta \rightarrow \infty$) or that it chooses more randomly ($\beta \rightarrow 0$). Parameter π is a stickiness parameter, which adds a certain amount of the value of π to the value estimate of the lastly chosen lever. In this case, positive values of π indicate a preference for the lastly chosen lever, negative values of π indicate a preference for the lastly unchosen lever, and π approaching 0 indicates that the side of the lastly chosen lever does not affect the next lever choice. ϕ is a boolean that was attributed the value 1 if that lever was chosen in the last trial (thus an amount of the value of π will be added to the value function), and 0 if that lever was not chosen in the last trial.

To obtain reliable model parameter estimates on a population level, we used maximum a posteriori estimation. In brief, we applied a prior distribution over the parameter values, and considered any new evidence from the animal's choice behavior to determine a posterior probability using Bayes' rule. These posterior probabilities were marginalized to get a point estimate of each session's best-fit parameter values. The used priors were: for α^{+} and α^{-} betapdf(1.5, 1.5); for π normpdf(0.5, 0.5); for β normpdf(2, 2).

All computational analyses were performed with Matlab R2014a (MathWorks Inc., USA).

Statistics

Statistical tests were performed in GraphPad Prism 6.0 (GraphPad Inc., USA). On all outcome parameters, a one-way repeated measures analysis of variance (one-way RM ANOVA) was performed, with estrous phase as a within-subjects repeated measures factor. This test was considered significant if $P < 0.05$, after which post-hoc Fisher's tests were performed. When data of more than one test per estrous phase was obtained (because data was collected from more than one cycle and/or animals were in a certain phase of the estrous cycle for more than one day), the outcome parameter values were averaged for these days. No statistical comparisons were made between males and females because the two groups were not tested in parallel and therefore equal testing conditions could not be ensured. In all graphs: **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, ^{ns} not significant.

Results

We observed a significant effect of estrous cycle on the total number of trials that the animals made during a session (Figure 1e). Animals that were in the estrus stage of the cycle made the lowest number of trials, and animals in the metestrus/diestrus stage the highest number of trials.

Performance in the task, measured as the total number of reversals that the animals achieved, revealed no significant differences between the three stages (Figure 1f). However, the total number of reversals is a compound measure for performance in the task, that does not necessarily inform about the underlying component processes. To gain insight into whether these underlying processes were modulated by the cycle, we fit the trial-by-trial data in the session to a computational reinforcement learning model[11], and used maximum a posteriori estimation[12] to determine the parameter values that best described the behavior of the animals (Figure 1d). After estimating the value of the four model parameters for each session, and comparing these between the different stages of the cycle (Figure 1g), we observed a significant decrease in reward learning parameter α+ during the proestrus stage, indicative of a lower impact of positive feedback (i.e., reinforcement) on behavior. We further found that the estimate of explore/exploit parameter β was significantly reduced during the estrus stage. No significant changes were observed on the value estimates of punishment learning parameter α- and perseveration parameter π. We replicated these findings by fitting the data to a less complex model that only includes α+, α- and β as free parameters (Supplementary Figure 1). Overall, the value estimates of the parameters in female animals were roughly similar to those observed in males (Figure 1g), except that male animals made more trials in the task (Figure 1e).

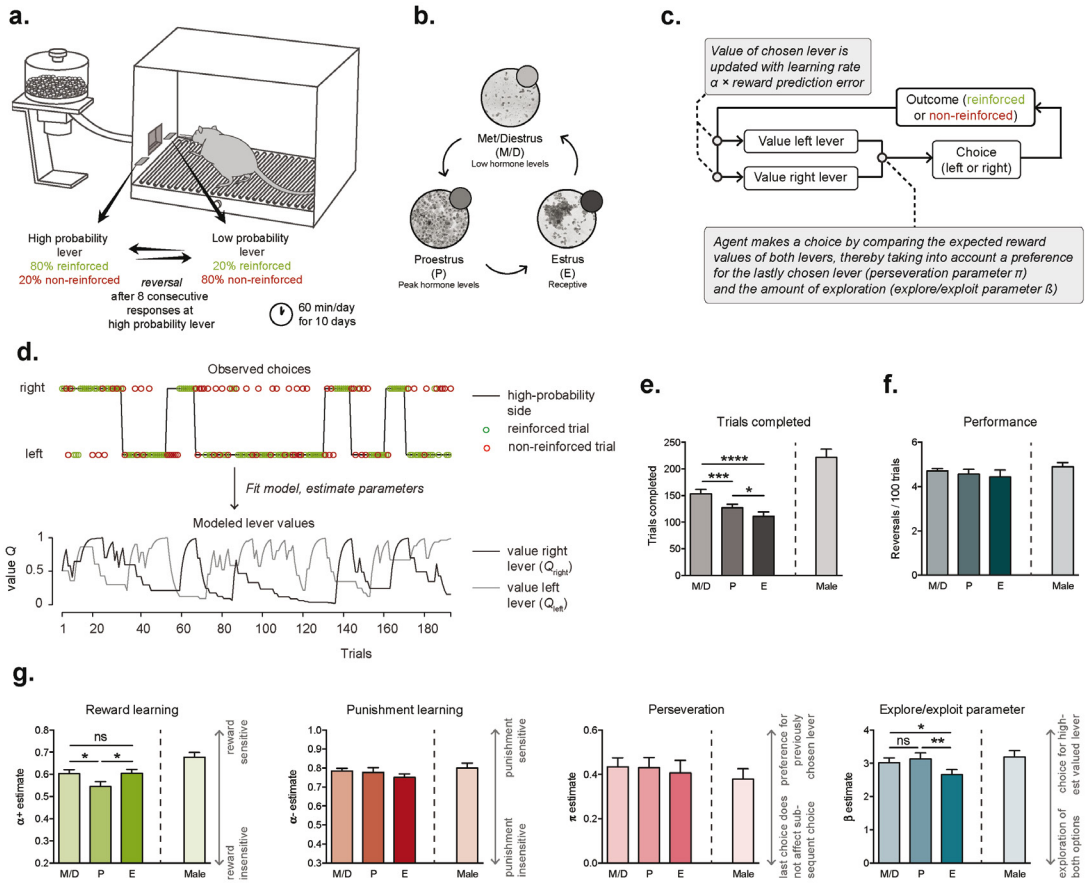


Figure 1

a. Probabilistic reversal learning setup. Hungry female animals could respond on two levers, one of which delivered sucrose reward with a high probability (80%, high-probability lever), and the other lever with a low probability (20%, low-probability lever). Every time the animal made eight consecutive responses on the high-probability lever, a reversal in reinforcement contingencies occurred, so that the previously low-probability lever became the high-probability lever, and vice versa. In this way, animals had to track the outcome of responding on each of the two levers over a series of trials and based hereon make a choice between them.

b. Example cytological images of samples from vaginal smears during the three stages of the estrous cycle.

c. Computational model.

d. Trial-to-trial data was fit to the computational model, and best-fit parameters were estimated.

e. Total trials completed by the female animals (n = 25) in the 60-minute session was significantly affected by the estrous cycle (Repeated measures ANOVA, $F_{2,48} = 21.22$, $P < 0.0001$). Post-hoc tests: **** $P < 0.0001$, *** $P = 0.0002$, * $P = 0.0188$. Male data (n = 18) is shown for illustrative purposes; these data were not included in the statistical analyses.

f. The total number of reversals was not affected by the cycle (ANOVA, $F_{2,48} = 0.48$, $P = 0.6209$).

g. Best-fit computational model parameters per estrous cycle stage. Reward learning: ANOVA $F_{2,48} = 3.995$, $P = 0.0248$; post-hoc tests met/diestrus (M/D) vs proestrus (P), $P = 0.0198$, M/D vs estrus (E), $P = 0.9425$, P vs E, $P = 0.0166$. Punishment learning: ANOVA $F_{2,48} = 1.637$, $P = 0.2052$. Perseveration: ANOVA $F_{2,48} = 0.1349$, $P = 0.8741$. Explore/exploit: ANOVA $F_{2,48} = 5.201$, $P = 0.0090$; post-hoc tests M/D vs P, $P = 0.4444$, M/D vs E, $P = 0.0243$, P vs E, $P = 0.0033$. Male data is shown for illustrative purposes.

Discussion

Our computational analyses reveal distinct changes in the processes underlying value-based decision making across the rat estrous cycle. The observed decrease in reward learning parameter *α+* during the proestrus stage is indicative of a lower impact of positive feedback (i.e., reinforcement) on behavior. This stage of the cycle is characterized by peak levels of the sex hormones progesterone and estradiol, and thus suggests a direct effect of gonadal steroids on reward processing, especially since reward learning was higher in the estrus stage of the cycle, when circulating hormone levels decline. This decreased focus on recent reward might also explain the reduction in trials completed, possibly reflecting attenuated motivation to obtain food reward (Supplementary Figure 2). However, the observed effect on motivation may also be the result of cyclic changes in appetite[13].

The reduction in the value estimate of explore/exploit parameter *β* during estrus indicates that sexually receptive females chose more stochastically (i.e., shifting from exploitation to exploration of the response options) than during the non-receptive stages of the cycle, perhaps reflecting a general increase in exploratory behavior. At the same time, this increase in exploration may have resulted in reduced task engagement, leading to a decrease in the number of trials completed (Supplementary Figure 2). Whether such cyclic changes in exploration have some evolutionary advantage, for example by promoting search for a sexual partner, remains to be investigated.

Researchers are increasingly encouraged to include female animals in preclinical experiments, with the aim to increase the translational value of animal research. In this regard, our data provide further insight into the complexity of value-based decision making and its sex-specific modulation. Importantly, behavioral data from intact female animals should be properly controlled for the estrous cycle, since many behavioral tasks in neuroscience involve (food) reward, and are therefore subject to changes in value-based learning, motivation and appetite.

In sum, we provide direct evidence that reward learning, exploration and motivation, but not punishment learning and perseveration, fluctuate during the estrous cycle in female rats. Although cyclic changes in value-based decision making have been observed before, which computational components underlie these changes had not yet been elucidated. It is well known that gonadal steroids have widespread effects on the brain, including the mesocorticolimbic dopamine system[14], which is an important hub for value-based learning[5]. It is therefore likely that estradiol and progesterone affect reinforcement learning through corticolimbic mechanisms, to promote adaptive survival-directed behavior in females.

ACKNOWLEDGEMENTS

Funding was provided by the European Union Seventh Framework Programme (grant agreement number 607310; Nudge-It), and by the Consortium on Individual Development (CID), which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003).

AUTHOR CONTRIBUTIONS

J.P.H.V. performed the behavioral experiments and analyzed the data. J.K. obtained the vaginal smears and determined the estrous cycle stage. L.J.M.J. and R.A.H.A supervised the experiments. All authors wrote the manuscript and have approved the final version of this paper.

CONFLICT OF INTEREST

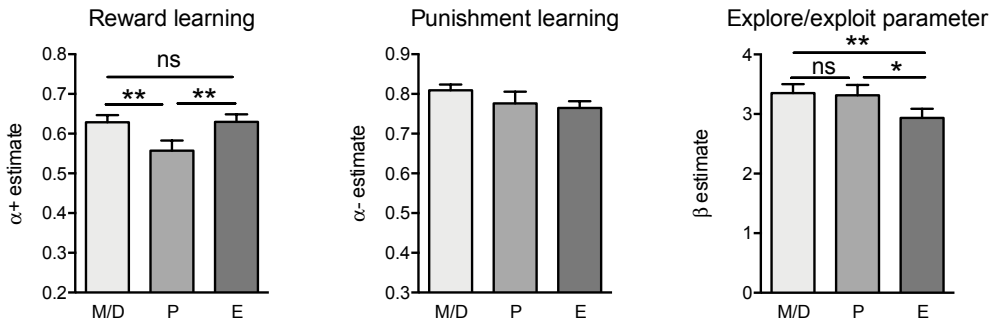
The authors declare that they have no conflict of interest.

DATA AVAILABILITY
All data is publicly available at github.com/jeroenphv/EstrousCycle.

References

1. Sutton, R. S. & Barto, A. G. Reinforcement learning: An introduction. (MIT press, 1998).
2. Maia, T. V. & Frank, M. J. From reinforcement learning models to psychiatric and neurological disorders. Nature neuroscience 14, 154-162 (2011).
3. Hendrick, V., Altshuler, L. L. & Burt, V. K. Course of Psychiatric Disorders across the Menstrual Cycle. Harvard Review of Psychiatry 4, 200-207 (1996).
4. Bari, A, Theobald, D.E., Caprioli, D., Mar, A. C., Aidoo-Micah, A. & Dalley, J.W. Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. Neuropsychopharmacology 35, 1290-1301 (2010).
5. Verharen, J. P. H., de Jong, J. W., Roelofs, T. J., Huffels, C. F. M., van Zessen, R., Luijendijk, M. C., Hamelink, R., Willuhn, I., den Ouden, H. E., van der Plasse, G., Adan, R. A. H. & Vanderschuren, L. J. M. J. A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. Nature communications 9 (2018).
6. Goldman, J.M., Murr, A.S. & Cooper, R.L. The rodent estrous cycle: characterization of vaginal cytology and its utility in toxicological studies. Birth Defects Research Part B: Developmental and Reproductive Toxicology 80, 84-97 (2007).
7. Cora, M. C., Kooistra, L. & Travlos, G. Vaginal Cytology of the Laboratory Rat and Mouse: Review and Criteria for the Staging of the Estrous Cycle Using Stained Vaginal Smears. Toxicol Pathol 43, 776-793 (2015).
8. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 2, 64-99 (1972).
9. Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A. & Daw, N. D. Differential roles of human striatum and amygdala in associative learning. Nature neuroscience 14, 1250-1252 (2011).
10. Posch, M. Win–Stay, Lose–Shift Strategies for Repeated Games—Memory Length, Aspiration Levels and Noise. J. Theor. Biol. 198, 183-195 (1999).
11. Gershman, S. J. Empirical priors for reinforcement learning models. Journal of Mathematical Psychology 71, 1-6 (2016).
12. Daw, N. in Decision Making, Affect, and Learning: Attention and Performance XXIII Ch. 6, (2009).
13. Tarttelin, M. F. & Gorski, R. A. Variations in food and water intake in the normal and acyclic female rat. Physiology & Behavior 7, 847-852 (1971).
14. McEwen, B. S. & Alves, S. E. Estrogen Actions in the Central Nervous System. Endocrine Reviews 20, 279-307 (1999).

SUPPLEMENTARY FIGURE 1



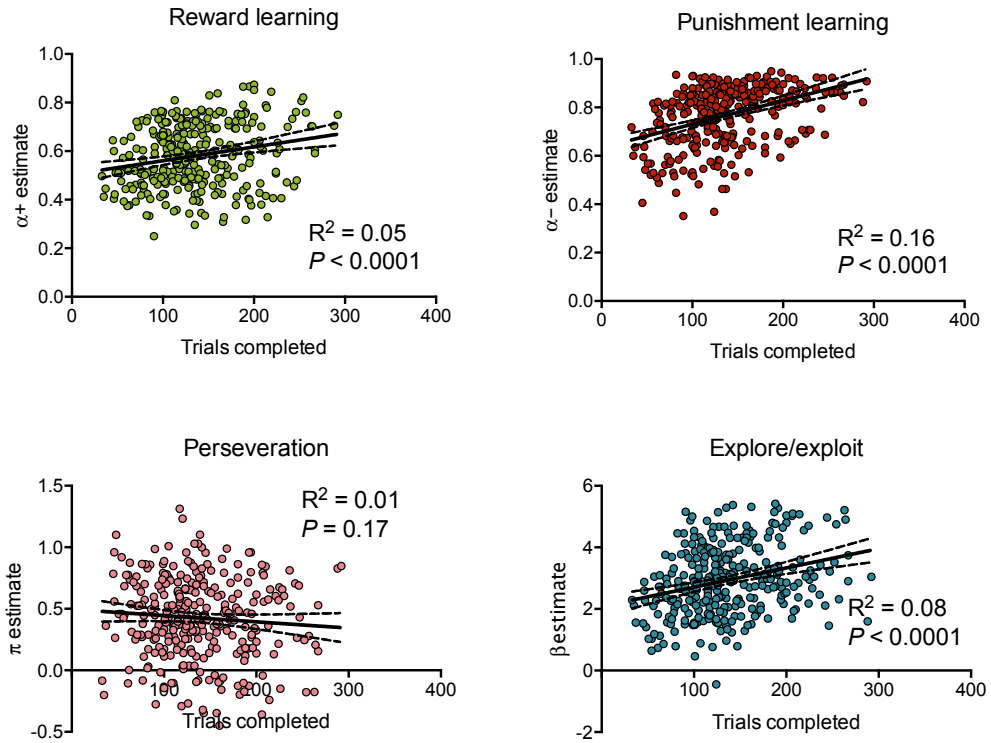
Fit of the data to a less complex model that does not include perseveration parameter π replicated the main findings of this paper. A significant reduction in reward learning parameter α+ was observed during proestrus, and a significant reduction in explore/exploit parameter β was observed during estrus.

Reward learning
One-way repeated measures ANOVA: $F_{2,48} = 5.128$, P = 0.0096 **
Post-hoc M/D vs P: $t_{48} = 2.747$, P = 0.0084 **
Post-hoc M/D vs E: $t_{48} = 0.052$, P = 0.9589
Post-hoc P vs E: $t_{48} = 2.799$, P = 0.0074 **

Punishment learning
One-way repeated measures ANOVA: $F_{2,48} = 2.222$, P = 0.1194

Explore/exploit
One-way repeated measures ANOVA: $F_{2,48} = 4.789$, P = 0.0127 *
Post-hoc M/D vs P: $t_{48} = 0.261$, P = 0.7953
Post-hoc M/D vs E: $t_{48} = 2.801$, P = 0.0073 **
Post-hoc P vs E: $t_{48} = 2.540$, P = 0.0144 *

SUPPLEMENTARY FIGURE 2



The total trials completed in the task positively correlated with the value estimates of reward learning parameter α+, punishment learning parameter α+ and explore/exploit β, but not with perseveration parameter π. n = 300 sessions from 30 rats. R^2 and P represent the variance explained and statistical significance of the linear regression analysis.

SUPPLEMENTARY TABLE 1

	Model	Free parameters	Aggregate LL	Aggregate AIC	# of sessions best described by model	XP	PXP
1	Rescorla-Wagner 1	α, β	-17800.7	36801.3	42 of 300	0	0
2	Rescorla-Wagner 2	α+, α-, β	-17289.9	36379.7	46 of 300	0	0
3	**Rescorla-Wagner 3**	**α+, α-, π, β**	**-16598.9**	**35597.8**	**126 of 300**	**1**	**1**
4	Pearce-Hall	η, β	-18581.3	38362.5	37 of 300	0	0
5	Rescorla-Wagner-Pearce-Hall hybrid	α, η, β	-17874.9	37549.9	36 of 300	0	0
6	Win-stay, lose-switch	β	-22124.8	44849.7	12 of 300	0	0
7	Random choice model	-	-28373.3	56746.6	1 of 300	0	0

Model comparisons, based on n = 300 sessions (10 sessions × 30 rats; total 40,934 trials).

Parameters: α, general Rescorla-Wagner learning rate; α+, Rescorla-Wagner reward learning rate; α-, Rescorla-Wagner punishment learning rate; π, stickiness or perseveration parameter; η, Pearce-Hall associability factor; β, choice stochasticity parameter (i.e., Softmax inverse temperature or explore/exploit parameter).

Abbreviations: LL, log-likelihood; AIC, Akaike Information Criterion; XP, exceedance probability; PXP, protected exceedance probability.

CHAPTER 6

Corticolimbic mechanisms of behavioral inhibition under threat of punishment

Jeroen P.H. Verharen
Mauri van den Heuvel
Mieneke C.M. Luijendijk
Louk J.M.J. Vanderschuren*
Roger A.H. Adan*

* Equal contribution

Manuscript under review

Highlights

- We developed a novel behavioral task that studies behavioral inhibition when animals need to balance reward pursuit and punishment avoidance
- Inactivation of the medial prefrontal cortex reduces inhibitory control
- The ventral striatum and basolateral amygdala are important for task engagement and threat cue processing, respectively

Techniques

Behavioral pharmacology

CHAPTER 6

Being able to limit the pursuit of reward in order to prevent negative consequences is an important expression of behavioral inhibition. Everyday examples of an inability to exert such control over behavior are the overconsumption of food and drugs of abuse, which are important factors in the development of obesity and addiction, respectively. Here, we use a behavioral task that assesses the ability of rats to exert behavioral restraint at the mere sight of palatable food during the presentation of an audiovisual threat cue to investigate the corticolimbic underpinnings of behavioral inhibition. We demonstrate a prominent role for the medial prefrontal cortex in the exertion of control over behavior under threat of punishment. Moreover, task engagement relies on function of the ventral striatum, whereas the basolateral amygdala mediates processing of a threat cue. Together, these data show that inhibition of reward pursuit requires the coordinated action of a network of corticolimbic structures.

Introduction

In a world where food is abundantly available, it can be hard to resist the temptation to eat highly palatable, yet unhealthy foods, while being aware of the negative health consequences this may have. As such, a healthy lifestyle requires one to control the urge to eat tasty foods. This can be especially challenging during dieting, when the body is in a negative energy state, and food cues are more salient than usual[1]. Accordingly, reduced behavioral inhibition has shown to be an important factor in the development and maintenance of overweight in children[2] and adults[3].

Besides its role in eating and dieting, deficiencies in inhibitory control have been implicated in a wide variety of maladaptive behaviors, ranging from failures in everyday life, like an inability to attain goals, to mental disorders, like substance addiction, attention-deficit/hyperactivity disorder (ADHD) and obsessive-compulsive disorder[4-6]. Behavioral inhibition is generally assumed to be a multifaceted phenomenon, whereby a distinction can be made between control over actions and control over choices and decisions[6,7]. These processes have been widely studied using laboratory tasks of impulsivity, which have tremendously progressed our understanding of the neural circuits involved in behavioral control[6-8]. Indeed, the distinction between choices, decision and actions is theoretically and mechanistically useful, but many everyday cases in which control over behavior is compromised consists of a combination of these processes. For example, an inability to resist a tasty dessert during dieting can sometimes be initiated by a thoughtless walk to the fridge, but during consumption, many decision moments take place in which one can reflect on his or her behavior and consider the consequences of continued eating in the short and long term.

In an attempt to capture behavioral inhibition in an ecologically valid fashion, we have developed a behavioral task in rats that measures the ability of the animals to inhibit the urge to consume a highly palatable food reward when a stimulus is presented that signals that sugar retrieval will be punished with a mild electric foot shock. Such a threat puts the animals in a conflict situation, in which a natural approach response to food competes with the natural avoidance response to danger. As such, our task assesses inhibitory control over an innately present desire.

Here, we investigated the corticolimbic substrates of behavioral control. Adaptive inhibition of behavior is thought to rely on functional activity in a network of regions including the prefrontal cortex (PFC), ventral striatum and amygdala, that has been implicated in the processing of emotionally relevant cues, the selection of appropriate behavioral strategies and the transmission of such strategies into goal-directed behavior[8-11]. Therefore, we tested how pharmacological inactivation of these brain structures altered behavior in this task. We hypothesized that inactivation of these structures would lead to marked, but behaviorally dissociable impairments in task performance.

Materials and methods

Animals

A total of 121 male Long-Evans rats (Rj:Orl, Janvier labs, France), weighing 250-300g at the start of the experiment, were used for this study. Animals were kept on a 12h/12h reversed day-night cycle (lights off at 8 A.M.). Animals were socially housed before surgery, but singly after surgery to prevent damage to the head implant. Experimental procedures were approved by the Animal Ethics Committee of Utrecht University and the Dutch Central Animal Testing Committee and they were conducted in agreement with Dutch laws (Wet op de Dierproeven, 2014) and European guidelines (2010/63/EU).

Surgeries

For placement of the guide cannulas, animals were anaesthetized with an intramuscular injection of a mixture of 0.315 mg/kg fentanyl and 10 mg/kg fluanisone (Hypnorm, Janssen Pharmaceutica, Belgium), and placed in a stereotaxic apparatus (David Kopf Instruments, United States). An incision was made along the midline of the skull, and two small craniotomies were made bilaterally above the brain region of interest. The following coordinates were used for placement of the guide cannulas:

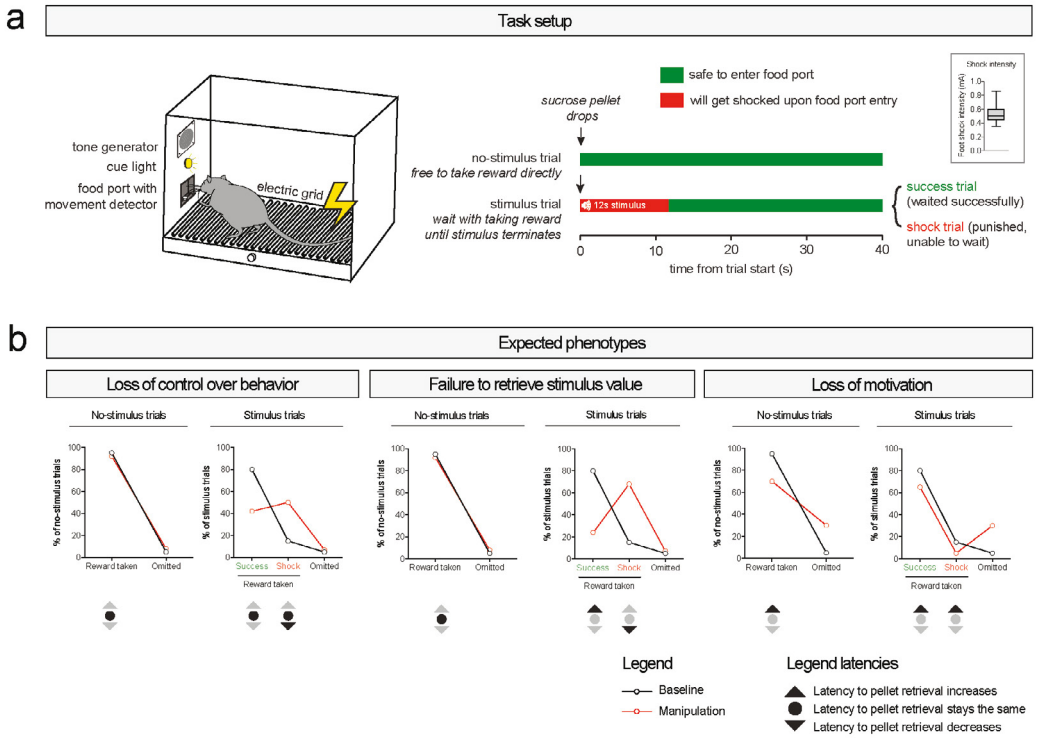
Prelimbic ctx	AP +3.2 mm	ML ±0.6 mm	DV -2.6 mm from skull
Infralimbic ctx	AP +3.2 mm	ML ±0.6 mm	DV -4.3 mm from skull
Medial orbitofrontal ctx	AP +4.4 mm	ML ±0.6 mm	DV -3.8 mm from skull
Anterior cingulate ctx	AP +2.0 mm	ML ±0.6 mm	DV -2.2 mm from skull
Lateral orbitofrontal ctx	AP +3.6 mm	ML ±2.6 mm	DV -3.7 mm from skull under 5° angle
Basolateral amygdala	AP -3.0 mm	ML ±5.0 mm	DV -7.5 mm from skull
Ventral striatum (core)	AP +1.2 mm	ML ±2.1 mm	DV -6.3 mm from skull under 5° angle
Ventral striatum (shell)	AP +1.2 mm	ML ±2.7 mm	DV -7.0 mm from skull under 10° angle
Dorsomedial striatum	AP +1.2 mm	ML ±2.3 mm	DV -4.1 mm from skull under 5° angle
Dorsolateral striatum	AP +1.2 mm	ML ±3.4 mm	DV -4.1 mm from skull
Olfactory cortex	AP +3.6 mm	ML ±2.2 mm	DV -4.4 mm from skull

For the brain regions of the medial PFC (prelimbic, infralimbic, medial orbitofrontal, and anterior cingulate cortices), 23G bilateral guide cannulas were used that had a double protrusion, spaced 1.2 mm apart (Plastics One, United States). For the other regions (lateral orbitofrontal cortex, basolateral amygdala, olfactory cortex, striatum), two 23G single guide cannulas (Plastics One, United States) were placed bilaterally.

Guide cannulas were lowered to the desired coordinates, secured with screws, dental glue (C&B Metabond, Parkell Prod Inc., United States) and dental cement, and the skin around the cemented cap was sutured. Dummy cannulas were placed inside the guide cannulas. Post-surgery, the animals were injected with 5 mg/kg carprofen for pain relief (1x/day, for 3 days, subcutaneously) and saline for rehydration (10 ml once, subcutaneously), and they were allowed to recover for 7 days before behavioral training continued.

Experimental procedures

Behavioral testing took place during the dark phase of the reversed 12h/12h day-night cycle. The task was conducted in operant conditioning chambers (31 x 24 x 21 cm; MedPC, Med Associates Inc., United States), placed in sound-attenuating cubicles. The chamber contained a shock grid floor, a 28V/100mA houselight, and in the right wall a food port with infrared movement detection, two 28V/100mA cue lights (flanking the food port), and a tone



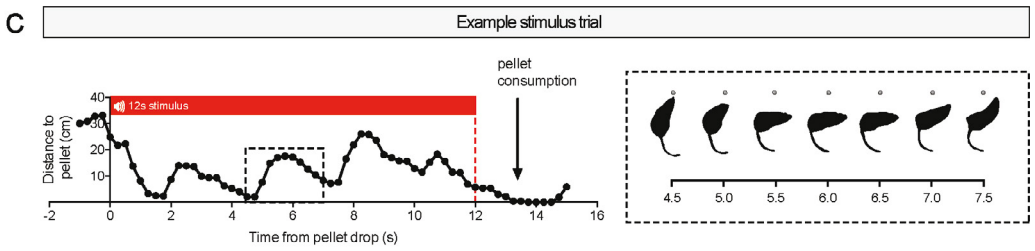


Figure 1
a. Behavioral setup. The task comprised 60 trials in which a sucrose pellet was delivered into a food port. In half of the trials, animals could take the pellet directly without any negative consequences ('no-stimulus trials'). In the other half of the trials, pellet delivery was accompanied by a 12s audiovisual stimulus, that signaled to the animals that they had to wait with entering the food port until stimulus termination ('stimulus trials'). Food port entry during the stimulus was detected by an infra-red movement detector and was punished with a 0.3s electric foot shock. Inset shows the individual animals' foot shock intensities (median ± 25th-75th percentile, whiskers extend to minimum and maximum values).
b, Possible phenotypes after (neural) manipulation. Note that for no-stimulus trials, both options ('Reward taken' and 'Omitted') add up to 100%, as well as for the options during the stimulus trials ('Reward taken - Success', 'Reward taken - Shock', 'Omitted'). Dark arrows under graphs represent possible changes in latency of pellet retrieval for each trial type.
c. Quantification of a trial from Supplementary movie 2, demonstrating 'attract and repel' behavior directed towards and away from the food receptacle during a stimulus trial.

generator (4500 Hz). A pellet dispenser delivered 45 mg sucrose pellets into the food port (SP; 5TUL; TestDiet, United States). Operant chambers were controlled by MedPC software (Version IV; Med Associates Inc., United States). Animals were kept on food restriction during the training phase (~4 gram chow per 100 gram body weight) and always had *ad libitum* access to water in their home cage. After successful training, animals received *ad libitum* chow. However, before behavioral testing, animals were food restricted for ~3 hours.

Task
A session consisted of 60 trials of 40 seconds each. At the start of every trial, one sucrose pellet was delivered into the food port, regardless of trial type (Fig. 1a). The trials were pseudorandomly distributed so that 30 trials were assigned as 'no-stimulus trials', and the remaining 30 trials were assigned as 'stimulus trials'. This order of trials was the same for all the animals, so that a larger cohort of animals could be tested simultaneously in the same room, without leakage of stimulus sound between the boxes. The house light was illuminated for the entire length of the session.

All trials started with the delivery of a sucrose pellet into the food port. During no-stimulus trials, the animals were allowed to enter the food port (i.e., consume the pellet) directly, which was detected by disruption of the infrared photobeam in the port. During stimulus trials, pellet delivery co-incided with the onset of a continuous tone and cue light stimulus, which lasted for 12 seconds, functioning as a threat signal to the animal. That is, the tone and light cue indicated that the animals had to wait with food port entry (and pellet consumption) until stimulus termination. If the animal managed to wait for 12 seconds, it could freely enter the food port and consume the sugar without scheduled consequences; this was called a 'success' trial. Food port entry during the stimulus, however, terminated the stimulus and delivered a 0.3 second foot shock to the animal; this was termed a 'shock' trial. The intensity of this foot shock was determined during the training phase for each animal individually, but it was kept constant for each animal during the experiment (median foot shock intensity 0.50 mA; see also Fig. 1a).

During the task, MedPC software recorded, for each trial, the type of trial (stimulus or no-stimulus), the response of the animal (pellet retrieved or not, and for stimulus trials if the trial was punished or not), the timestamp of the pellet drop, and the timestamp of the response of the animal. Since latencies of pellet retrieval were usually not normally distributed within a session, the median latency for each trial type per session, per animal was used in the analysis.

When the animal did not enter the food port (and consume the pellet) during a trial (i.e., within 40 seconds), it was regarded as an omission, and this prevented further pellet delivery (and hence pellet accumulation in the food port) until the next food port entry. To control for these omissions, we computed a shock index, which is the number of shock trials as a fraction of the number of shock+success trials. In other words, this index is a measure for the amount of stimulus trials during which the animal entered the food port during stimulus presentation, corrected for the number of omissions, and thus represents a measure of (loss of) control over behavior.

Expected phenotypes
Based on the trial outcomes and the speed with which animals retrieve the pellets, different behavioral phenotypes can be discerned (Fig. 1b). First, impaired inhibition of behavior, in which animals are not able to refrain from taking the sucrose pellet for the entire stimulus period, would be characterized by an increase in shock trials, at the expense of the number of success trials (Fig. 1b, left panel). Latency of pellet retrieval during shock trials is likely to be decreased compared to control conditions, i.e., *if* animals show reduced control over behavior, this may happen earlier in the stimulus period. Behavior during no-stimulus trials should be unchanged, and neither should be the latency of pellet retrieval during success

trials (i.e., the speed of food port entry after stimulus offset).

Second, when the animal's capability of retrieving the value of the stimulus is compromised, animals would behave as if there was no threat signal presented at all (Fig. 1b, middle panel). This would lead to a similar behavioral pattern as after loss of control over behavior, but with different latency effects. During shock trials, in which the animals take the pellet during the stimulus, retrieval latency should be shorter, as animals are less able to distinguish between no-stimulus and stimulus trials. Similarly, latency of food port entry after stimulus offset, during success trials, is likely to be higher, since animals will not successfully retrieve the termination of the threat signal.

Third, a loss of motivation to obtain reward would increase the amount of omissions, both in no-stimulus, as well as in stimulus trials (Fig. 1b, right panel). The number of shock trials will likely be low, as it will be easier for the animals to wait with pellet retrieval until after stimulus termination. Furthermore, latency until pellet retrieval is likely to be increased in all trials.

Finally, behavior could be disrupted by a combination of these three phenotypes, which could lead to a variety of patterns in trial outcomes and latencies.

Task training

Animals were trained once or twice a day, for 5-7 days per week, starting with magazine training, which was the same task as described above except that exclusively no-stimulus trials were presented. Thus, 60 sucrose pellets were delivered into the food port with an interval of 40 seconds. If the animals made less than 5 omissions in a session, training progressed to the final training phase (see Fig. S1), which was the task version described above.

In the first session of the final training phase, foot shock intensity was set to 0.35 mA. If more than half of the stimulus trials were punished, it was assumed that the intensity was too low to induce effective punishment, hence the foot shock intensity of the next session was increased with 0.05 or 0.1 mA. Similarly, if an animal made many omissions, it was assumed that the foot shock was too intense, and shock intensity was decreased with 0.05 mA in the next session. After animals reached the criterion of 20 success trials out of 30 stimulus trials (meaning that the rat waited with pellet retrieval in 2/3 of the stimulus trials), foot shock intensity was kept constant for the remainder of the experiment. All animals learned the task, so no 'non-learners' had to be excluded from the experiment.

Infusions

For the intracranial infusions into the bilateral guide cannulas, double injectors were used that protruded 1 mm beyond the end of the guides. For the single guide cannulas, injectors were used that protruded ~0.4 mm beyond the end of the guide. Animals were habituated to the procedure the day before the experiment, by an infusion of 0.3 µl saline through the cannulas.

On testing day, animals received an infusion of a cocktail of baclofen (1 nmol; Sigma-Aldrich, Netherlands) and muscimol (0.1 nmol; Sigma-Aldrich, Netherlands) dissolved in 0.3 µl saline[12], or 0.3 µl saline as a control (counterbalanced between days, 24h apart) using a syringe pump (Harvard Apparatus, United States) set at an infusion rate of 0.5 µl/min. After infusion, the injectors were kept in place for an additional 30 seconds to allow the drug to properly diffuse into the tissue. After infusion, the dummy cannulas were placed back into the guides, and the animals were returned to their home cage for 10-20 minutes, before experimental testing commenced.

Free-feeding assay

In the free feeding assay, animals were infused with baclofen/muscimol or saline, and placed back into their home cage for 2 hours. Animals had *ad libitum* access to chow in a feeding rack that was attached to the wall of the home cage. Food was weighed at the beginning of the experiment and again two hours later. Animals were measured twice, once after infusion of baclofen/muscimol and once after infusion of saline (counterbalanced between subjects; 24h apart).

Tail withdrawal test

The tail withdrawal test (adapted from refs. 13 and 14) took place during the 2-hour free feeding assay (which occurred twice; treatment counterbalanced between subjects, see above). In this task, the animal was fixated with a towel, and 3-5 cm of the animal's tail was placed in a beaker containing water of 50 (± 1) °C. The test was filmed, and latency until tail withdrawal was scored from the movies in a frame-by-frame manner, by a researcher blind to the treatment (baclofen/muscimol or saline).

Histological verification

After the behavioral experiments, animals were transcardially perfused with phosphate-buffered saline followed by 4% paraformaldehyde in phosphate-buffered saline. Brains were post-fixed in 4% paraformaldehyde in phosphate-buffered saline for 24 hours at 4°C followed by a 30% sucrose solution at 4°C. Next, brains were cut in coronal slices of 50 µm using a cryostat. Brain slices were mounted and colored with 5% Giemsa (Sigma-Aldrich, The Netherlands) dissolved in distilled water. Infusion sites were histologically verified by a researcher blind to the experimental results.

Exclusion criteria

Five animals were excluded based on misplacement of the cannulas: infralimbic cortex, 1; anterior cingulate cortex, 1; medial orbitofrontal cortex, 1; lateral orbitofrontal cortex, 1; dorsomedial striatum, 1. Ten animals died during surgery: dorsomedial striatum, 1; infralimbic cortex, 2; anterior cingulate cortex, 1; medial orbitofrontal cortex, 2; lateral orbitofrontal cortex, 2; basolateral amygdala, 1; dorsomedial striatum, 1. One animal was excluded from the basolateral amygdala group due to blockage of the cannula. Infusions into the ventral striatum were initially targeted separately at the nucleus accumbens shell versus core, but were later combined into one ventral striatum group, because the areas were difficult to histologically distinguish. One animal from this group was excluded because it lost its headcap. Data from one animal was removed from the ventral striatum infusion experiment, because the pellet dispenser did not work during the saline session.

Code availability

The MedPC script of the task is available at http://www.github.com/jeroenphv.

Statistics

Statistical tests were performed with Prism 6.0 (GraphPad Software Inc., United States). Statistical tests were a within-animal (i.e., paired) comparison, in which baclofen/muscimol treatment was compared to saline (baseline) treatment. In these experiments, brain region was not included as a between-subjects factor, because we expected different behavioral phenotypes after inactivation of the different brain regions. In the free-feeding assay and tail withdrawal test, a two-way repeated measure analysis of variance (ANOVA) was used, with baclofen/muscimol versus saline as a within-subjects repeated measures factor, and treatment group (brain area) as a between-subjects factor. In all figures, statistical significance is denoted with the following range: $^{\#}P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. Extended statistics are presented in the Supplementary statistics table.

Results

Task behavior

All rats learned the task, i.e., they managed to wait to eat the pellet during the stimulus in the majority of trials (Supplementary Movie 1 and 2). In 'success trials', the rats retrieved the pellet quickly after tone offset, with an average latency of ~2.5s (Fig. S2). In 'shock trials', i.e., trials in which animals retrieved the pellet during the stimulus and hence received punishment, latencies of pellet retrieval were usually higher than in no-stimulus trials (~5s compared to ~1.5s in no-stimulus trials; Fig. S2), as if the rats managed to control their behavior for a fraction of the stimulus period before they reached out for the pellet. Interestingly, animals typically exhibited 'attract and repel' behavior directed towards and away from the food receptacle during behavioral control (Fig. 1c and Supplementary Movie 1 and 2).

Prefeeding devaluation evokes a loss-of-motivation phenotype

As described in the 'Materials and methods', different phenotypes can be discerned on the basis of the trial outcomes and the speed with which animals retrieve the pellet (Fig. 1b). As proof-of-principle, we pre-fed the animals with sucrose pellets before the task, to evoke a loss-of-motivation phenotype. Indeed, this induced a pattern of effects (Fig. S3) that matched expectations (Fig 1b, right panel). That is, the number of omissions increased, animals showed increased control over behavior, and there was a trend towards increased latencies of pellet retrieval during no-stimulus trials and success trials. No effect was observed on the latency of pellet retrieval in shock trials, but note that this latency was only based on data from 6 animals after prefeeding, as the other 6 animals never retrieved the pellet during the stimulus.

Medial PFC inactivation impairs inhibition over behavior

To study the involvement of different regions of the corticolimbic system to behavior in our task, we pharmacologically inactivated different regions of this system by means of intracranial infusions of the GABA receptor agonists baclofen and muscimol[12]. Inactivation of the prelimbic cortex significantly increased the number of shock trials, which came at the expense of the number of success trials, without affecting behavior during no-stimulus trials, or the number of omissions (Fig. 2a). Consequently, the shock index, which measures the fraction of stimulus trials in which the animal retrieved the pellet during the stimulus, thus receiving foot shock, increased significantly. No significant effects on the speed with which the animals retrieved the pellet were observed (Fig. 2a). This pattern of effects matches the phenotype corresponding to loss of control over behavior (Fig. 1b), suggesting that inactivation of the prelimbic cortex impaired the ability of animals to inhibit their urge to approach the pellet, despite the presence of the threat signal. Inactivation of the infralimbic cortex yielded the same pattern of effects (Fig. 2b). That is, an increase in the number of shock trials, a decrease in success trials and an increased shock index, without a change in behavior during no-stimulus trials or an effect on any of the latency measures.

Medial orbitofrontal cortex inactivation also impaired control over behavior, as apparent by a significant increase in the number of shock (but not a decrease in success) trials and thereby an increase in the shock index, although this effect was numerically more modest than after inactivation of the prelimbic and infralimbic cortices (Fig. 2c). In addition, it significantly decreased the latency of pellet retrieval during shock trials, indicating that if the animals entered the food port during the stimulus, this happened during an earlier stage

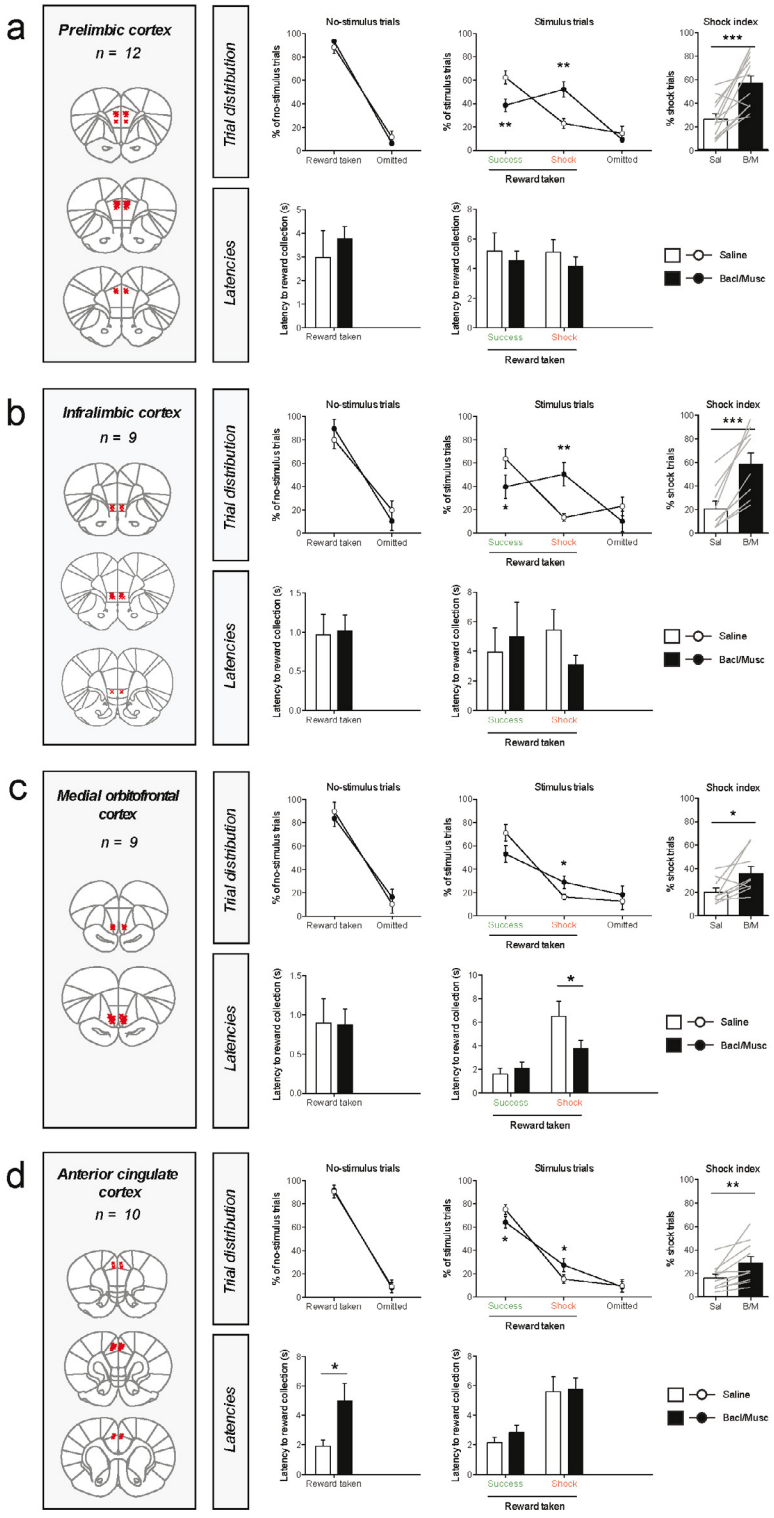


Figure 2: Effects of pharmacological inactivation of the medial PFC on task behavior. Red crosses in the coronal brain sections represent the infusion sites in each experiment. Gray lines in shock index graphs indicate individual animals. For latency analyses, the median latency per animal per trial type was used. * P < 0.05, ** P < 0.01, *** P < 0.001 in paired t-test; see Supplementary statistics table. →

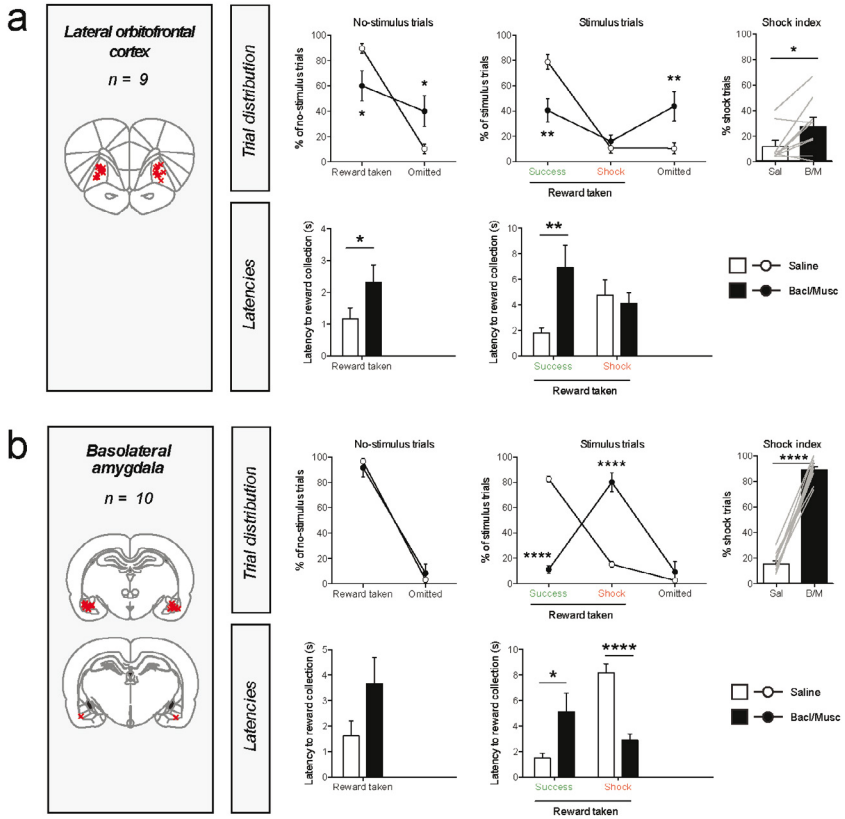


Figure 3: Pharmacological inactivation of lateral orbitofrontal cortex and basolateral amygdala
a. Effects of pharmacological inactivation of the lateral orbitofrontal cortex on task behavior.
b. Effects of pharmacological inactivation of the basolateral amygdala on task behavior.
Red crosses in the coronal brain sections represent the infusion sites in each experiment. Gray lines in shock index graphs indicate individual animals. * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001 in paired t-test; see Supplementary statistics table.

of stimulus presentation. This suggests that if the animals lost control over behavior after medial orbitofrontal cortex inactivation, they were able to inhibit themselves for a shorter period of time compared to baseline.

Inactivation of the anterior cingulate cortex also evoked disinhibition of behavior. Thus, the number of success trials was reduced, and the number of shock trials, as well as the shock index, increased. Latency analysis revealed that animals also became slower in pellet retrieval during no-stimulus trials, where they were allowed to consume the pellet directly without negative consequences (Fig. 2d). Thus, animals responded slower than under baseline conditions, suggesting that additional cognitive functions, such as attention, could be impaired.

Together, these data suggest that the prelimbic, infralimbic, and medial orbitofrontal cortices have a role in mediating control over behavior in this task, while the anterior cingulate cortex may also serve other cognitive functions that are necessary for correct task execution.

Lateral orbitofrontal cortex and basolateral amygdala inactivation disrupt task performance
Inactivation of the lateral orbitofrontal cortex and the basolateral amygdala impaired task performance, but in different ways. After inactivation of the lateral orbitofrontal cortex, we observed a significant increase in the number of omitted trials, as well as higher latencies of pellet retrieval in no-stimulus trials and success trials (Fig. 3a). Furthermore, no decrease in the number of shock trials was observed, hence the shock index was significantly increased. This pattern of effects suggests a reduction in task engagement.

Basolateral amygdala inactivation induced a dramatic increase in the number of shock trials, leading to a shock index of ~90%, meaning that 9 out of 10 port entries during stimulus trials were during the stimulus, and hence were punished (Fig. 3b). No effect was observed on the number of omissions during stimulus trials, and neither were there any effects on behavior during no-stimulus trials. Thus, basolateral amygdala inactivation only affected behavior around the stimulus presentation. Interestingly, the latency of pellet retrieval during shock trials was reduced (i.e., animals were able to control their behavior for a shorter amount of time), while an increase in latency was observed during success trials (i.e., after successful control, animals did not directly take the pellet after stimulus offset). This pattern of effects (see Fig. 1b) suggests that basolateral amygdala inactivation impaired the ability of the animals to retrieve the value of the stimulus, so that animals ostensibly behaved as if there was no threat signal presented.

Activity in the striatum is important for task engagement
After pharmacological inactivation of the ventral striatum, we observed a significant increase in the number of omissions during both trial types (Fig. 4a). During stimulus trials, this occurred at the expense of the total number of success trials. No change in the number of shock trials was observed. Because of this decrease in the number of success trials, we observed an increase in the shock index, as the relative amount of shock trials increased. Although no effects on latencies of pellet retrieval were found, it must be noted that due to the large amount of omissions, these latencies were based on a low number of trials (or even no trials for animals that made exclusively omissions). After histological verification of the infusion sites, we observed that most guide cannulas were positioned above the core region of the nucleus accumbens. However, when only analyzing the animals in which the infusions were targeted at the nucleus accumbens shell, the same pattern of effects was observed (Fig. S4).

Pharmacological inactivation of the dorsolateral and dorsomedial striatum showed a similar, although attenuated, pattern of effects. As such, dorsomedial striatum inactivation resulted in a trend towards an increase in the number of omissions during no-stimulus trials and a significant increase in omissions during stimulus trials, which was associated with a reduction in the number of success trials (Fig. 3b). After inactivation of the dorsolateral striatum, we observed a trend towards a reduction of the number of success trials and a significant increase in the shock index (Fig. 3c). No effect was observed on the shock index after inactivation of the dorsomedial striatum. Furthermore, a trend towards and a significant reduction in the latency of pellet retrieval during success trials was observed after pharmacological inactivation of the dorsolateral and dorsomedial striatum, respectively.

Control experiments
As a negative control region, we inactivated the dorsolateral part of the olfactory cortex, which is located ventral of the orbitofrontal cortex (Fig. 5a). This inactivation did not affect any of the behavioral parameters, indicating that the olfactory cortex is not essential for task performance, and it suggests that the infused GABA receptor agonists did not spread throughout the brain to induce infusion site-unspecific behavioral effects.

It is possible that changes in nociception underlie the effects we observed in this study. For example, animals may become less or more sensitive to the foot shock, resulting in

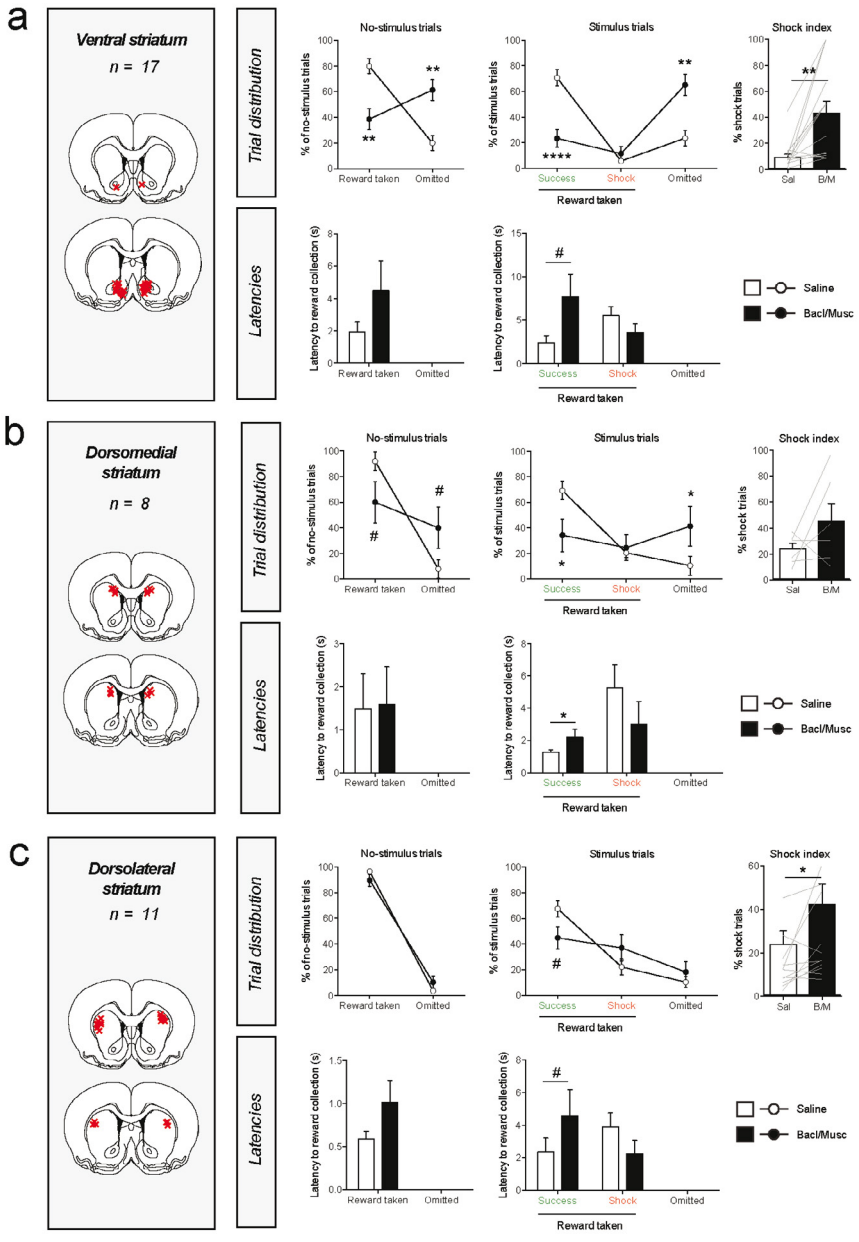


Figure 4: Effects of pharmacological inactivation of striatal subregions on task behavior. Red crosses in the coronal brain sections represent the infusion sites in each experiment. Gray lines in shock index graphs indicate individual animals.
**** P < 0.0001, *** P < 0.001, ** P < 0.01, * P < 0.05, # P < 0.1 in paired t-test (see Supplementary statistics table).

changes in behavior during stimulus trials. To control for this possible effect, we performed a tail withdrawal test in the animals, and observed no changes in latency to tail withdrawal after inactivation of the brain regions in which we found increases in the number of shock trials (Fig. 5b). We also conducted a free-feeding assay, since changes in appetite may change behavior in tasks that involve food reward (see Fig. S3). In none of the brain areas we observed changes in chow intake in the 2 hours following baclofen/muscimol infusion (Fig. 5c). These findings suggest that the observed effects in the behavioral task were not induced by changes in nociception or hunger.

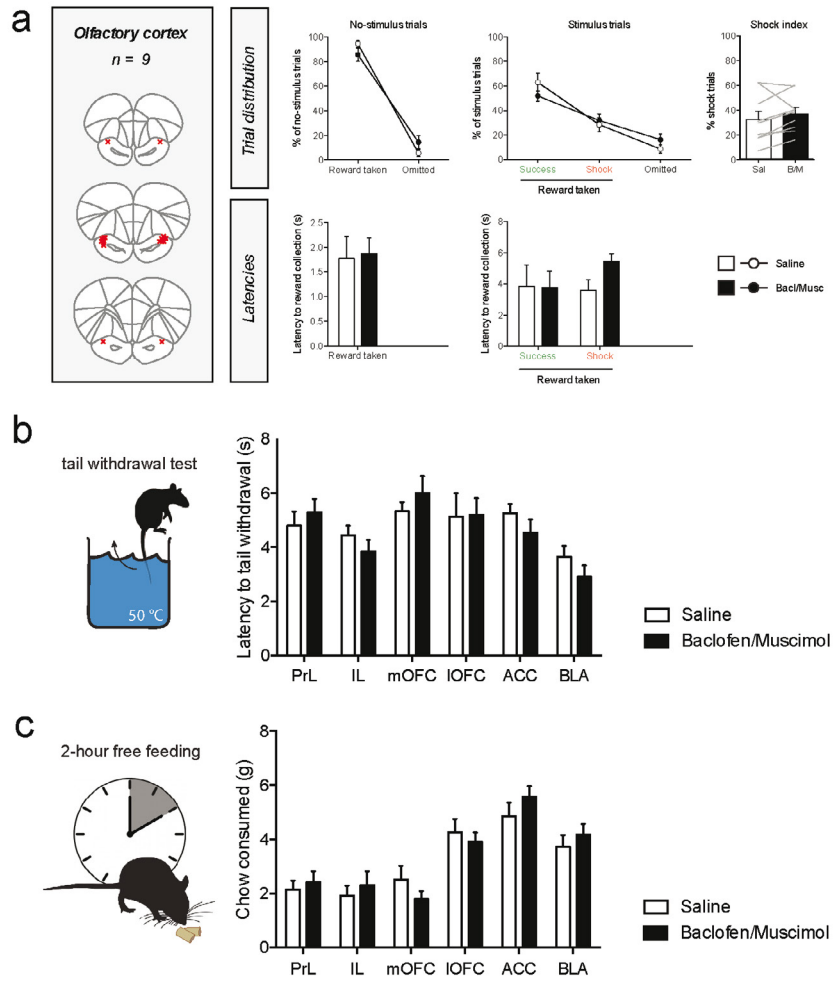


Figure 5: Control experiments.
a. Effects of pharmacological inactivation of the olfactory cortex on task behavior.
b. Pharmacological inactivations did not the change latency to tail withdrawal in a tail withdrawal test (2-way repeated measures ANOVA, main effect of infusion, and infusion × group interaction effect, both P > 0.05).
c. Pharmacological inactivations did not change chow intake in a free-feeding assay (2-way repeated measures ANOVA, main effect of infusion, and infusion × group interaction effect, both P > 0.05).
See Supplementary statistics table for statistics.

Discussion

In this study, we presented a novel task that studies the ability of rats to inhibit their urge to consume a visibly present food reward during the presentation of an audiovisual threat signal. Importantly, in this task, control over behavior comprises refraining from consumption of palatable food. Given that many day-to-day examples of loss of control over behavior encompass behavior directed at primary rewards, like food or drugs, this task aims to provide a more naturalistic approach to inhibition of behavior, since the animals have to balance two innate urges: approach to food reward versus an avoidance response to punishment. An additional benefit is that this task requires relatively little training (Fig. S1), as animals do not have to learn any operant responses to receive reward. Furthermore, using this task, we can discern different aspects of task behavior, including a failure to retrieve stimulus value, a lack of motivation, and compromised inhibition of behavior. One drawback of our task is that once animals have acquired the task, they often have very low baseline levels of shock trials, making strengthening of behavioral control hard to detect. Our approach shows similarities to certain models of relapse to drugs seeking, in which animals are confronted with behavioral conflict between pursuing (drug) reward and avoiding punishment[15,16].

Utilizing our new paradigm, we found that a wide array of corticolimbic regions is involved in the proper exertion of behavioral inhibition. Inactivation of the ventral parts of the medial PFC (prelimbic, infralimbic and medial orbitofrontal cortex) evoked a loss-of-control phenotype, i.e., a substantial increase in the number of shocks incurred, a decrease in success trials, without major changes in omissions or latencies (see Fig. 1b, left panel). This indicates that control over behavior under threat of punishment is governed by a neural network with the medial PFC as a core component, with a possible gradient across the dorsoventral axis. This is consistent with the notion that in our task, behavior is dependent on the weighing of the costs and benefits of a decision — a function that has been attributed to the PFC[17-19].

In contrast to the medial PFC, we observed a phenotype after pharmacological inactivation of the lateral orbitofrontal cortex that is reminiscent of a reduction in task engagement or incentive motivation. As such, we observed an increase in omissions, which came, during stimulus trials, at the expense of the number of success trials. Moreover, the significant increases in pellet retrieval latencies also point towards a reduction in motivation to obtain the reward or to a reduction in task engagement. Such a reduced task engagement may be the result of an inability to comprehend the task, consistent with recent theories of the lateral orbitofrontal cortex in guiding task execution by keeping a cognitive map of task structure[20,21], rather than the lateral obitofrontal cortex being directly involved in incentive motivation[9,22].

Inactivation of the basolateral amygdala evoked a phenotype that matched our hypothesized phenotype of a failure to retrieve stimulus value (Fig. 1b, middle panel). Thus, after infusion of baclofen and muscimol, we observed a dramatic increase in the number of shock trials, without effects on omissions or behavior during no-stimulus trials. Given that the latency of pellet retrieval in success trials was also increased, we speculate that animals did not comprehend the offset of the threat signal, suggesting that animals were not able to retrieve the value of the audiovisual stimulus. As such, the animals behaved as if no threat signal was presented during stimulus trials, and entered the food port during the stimulus, thus receiving foot shock punishment, on the vast majority of trials. This data is consistent with a wealth of literature that shows an involvement of the basolateral amygdala in responding to a conditioned cue[9,11,23,24], thereby evoking behavior that is ostensibly fearless and punishment insensitive.

After pharmacological inactivation of the ventral striatum, the animals behaved as if they were less motivated for the reward. As such, we observed an increased number of omissions, and a reduced number of rewards collected, even during no-stimulus trials. Inactivation of the dorsal parts of the striatum showed an attenuated version of this phenotype, suggesting that the motivational processes that are important for task performance are primarily encoded by ventral striatal circuits[10,25]. It must be noted, however, that we pharmacologically inactivated a relatively anterior part of the dorsal striatum, and there is evidence of a functional-anatomical gradient across its anteroposterior axis[26-28]. For example, goal-directed behavior is shown to be dependent on the poster, but not anterior, region of the dorsomedial striatum[29]. It might therefore be the case that behavioral control is mediated by the dorsal striatum, but that this process takes place in its posterior parts, especially because the absence of effects on the absolute number of shock trials challenges the classic view of the basal ganglia as part of the final common pathway of motoric Go/NoGo responses[30-32].

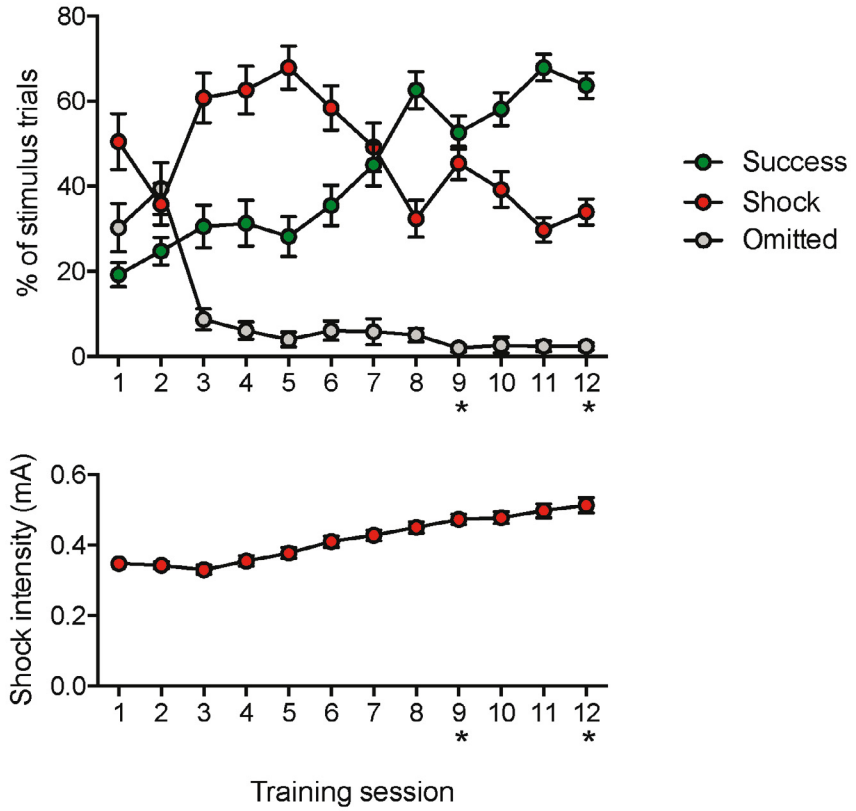
In sum, using a novel behavioral control task in rats, we show that behavioral inhibition is dependent on a network of corticolimbic areas, with the ventromedial PFC at its core, aided by striatal and orbitofrontal regions involved in task engagement and incentive motivation, and the basolateral amygdala to encode the value of relevant conditioned stimuli. Our data provide an important step in the dissection of the brain circuits involved in behavioral inhibition, and hence contribute to the understanding of behaviors that are associated with poor control over behavior, including binge eating and drug abuse.

References

1. van der Plasse, G. et al. Modulation of cue-induced firing of ventral tegmental area dopamine neurons by leptin and ghrelin. Int. J. Obes. (Lond) 39, 1742-1749 (2015).
2. Nederkoorn, C., Braet, C., Van Eijs, Y., Tanghe, A. & Jansen, A. Why obese children cannot resist food: the role of impulsivity. Eat Behav. 7, 315-322 (2006).
3. Nederkoorn, C., Houben, K., Hofmann, W., Roefs, A. & Jansen, A. Control yourself or just eat what you like? Weight gain over a year is predicted by an interactive effect of response inhibition and implicit preference for snack foods. Health Psychol. 29, 389-393 (2010).
4. Winstanley, C. A., Eagle, D. M. & Robbins, T. W. Behavioral models of impulsivity in relation to ADHD: translation between clinical and preclinical studies. Clin. Psychol. Rev. 26, 379-395 (2006).
5. Bari, A. & Robbins, T. W. Inhibition and impulsivity: behavioral and neural basis of response control. Prog. Neurobiol. 108, 44-79 (2013).
6. Dalley, J. W. & Robbins, T. W. Fractionating impulsivity: neuropsychiatric implications. Nat. rev. Neurosci. 18, 158-171 (2017).
7. Pattij, T. & Vanderschuren, L. J. The neuropharmacology of impulsive behaviour. Trends Pharmacol. Sci. 29, 192-199 (2008).
8. Dalley, J. W., Everitt, B. J. & Robbins, T. W. Impulsivity, compulsivity, and top-down cognitive control. Neuron 69, 680-694 (2011).
9. Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci. Biobehav. Rev. 26, 321-352 (2002).
10. Floresco, S. B. The nucleus accumbens: an interface between cognition, emotion, and action. Annu. Rev. Psychol. 66, 25-52 (2015).
11. Janak, P. H. & Tye, K. M. From circuits to behaviour in the amygdala. Nature 517, 284-292 (2015).
12. McFarland, K. & Kalivas, P. W. The circuitry mediating cocaine-induced reinstatement of drug-seeking behavior. J. Neurosci. 21, 8655-8663 (2001).
13. Nieh, E. H. et al. Decoding Neural Circuits that Control Compulsive Sucrose Seeking. Cell 160, 528-541 (2015).
14. Verharen, J. P. H. et al. A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. Nat. commun. 9 (2018).
15. Cooper, A., Barnea-Ygael, N., Levy, D., Shaham, Y. & Zangen, A. A conflict rat model of

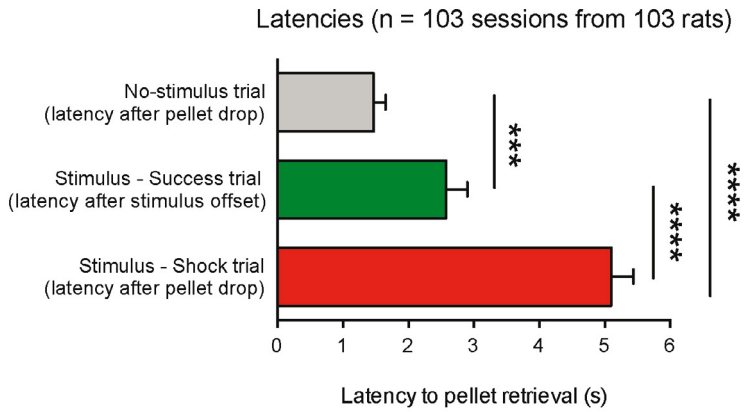
cue-induced relapse to cocaine seeking. Psychopharmacology (Berl) 194, 117-125 (2007).
16. Marchant, N. J., Khuc, T. N., Pickens, C. L., Bonci, A. & Shaham, Y. Context-induced relapse to alcohol seeking after punishment in a rat model. Biol. Psych. 73, 256-262 (2013).
17. Miller, E. K. & Cohen, J. D. An Integrative Theory of Prefrontal Cortex Function. Annu. Rev. Neurosci. 24, 167-202 (2001).
18. Floresco, S. B. Prefrontal dopamine and behavioral flexibility: shifting from an "inverted-U" toward a family of functions. Front Neurosci. 7, 62 (2013).
19. Tang, H., Sun, X., Li, B. & Luo, F. Neural representation of cost-benefit selections in medial prefrontal cortex of rats. Neurosci. letters 660, 115-121 (2017).
20. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal cortex as a cognitive map of task space. Neuron 81, 267-279 (2014).
21. Stalnaker, T. A., Cooch, N. K. & Schoenbaum, G. What the orbitofrontal cortex does not do. Nat. neurosci. 18, 620-627 (2015).
22. Izquierdo, A. Functional Heterogeneity within Rat Orbitofrontal Cortex in Reward Learning and Decision Making. J. Neurosci. 37, 10529-10540 (2017).
23. Davis, M. Neurobiology of fear responses: the role of the amygdala. The J. of neuropsych. and clin. neurosci. (1997).
24. Barad, M., Gean, P. & Lutz, B. The Role of the Amygdala in the Extinction of Conditioned Fear. Biol. Psych. 60, 322-328 (2006).
25. Voorn, P., Vanderschuren, L. J., Groenewegen, H. J., Robbins, T. W. & Pennartz, C. M. Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci. 27, 468-474 (2004).
26. Reynolds, S. M. & Berridge, K. C. Fear and feeding in the nucleus accumbens shell: rostrocaudal segregation of GABA-elicited defensive behavior versus eating behavior. J. Neurosci. 21, 3261-3270 (2001).
27. Pan, W., Mao, T. & Dudman, J. Inputs to the Dorsal Striatum of the Mouse Reflect the Parallel Circuit Architecture of the Forebrain. Frontiers in Neuroanat. 4 (2010).
28. Mestres-Missé, A., Turner, R. & Friederici, A. D. An anterior–posterior gradient of cognitive control within the dorsomedial striatum. NeuroImage 62, 41-47 (2012).
29. Yin, H. H., Ostlund, S. B., Knowlton, B. J. & Balleine, B. W. The role of the dorsomedial striatum in instrumental conditioning. Eur. J. of Neurosci. 22, 513-523 (2005).
30. Aron, A. R. et al. Converging evidence for a fronto-basal-ganglia network for inhibitory control of action and cognition. J. Neurosci. 27, 11860-11864 (2007).
31. Peters, J., LaLumiere, R. T. & Kalivas, P. W. Infralimbic prefrontal cortex is responsible for inhibiting cocaine seeking in extinguished rats. J. Neurosci. 28, 6046-6053.
32. Humphries, M. D. & Prescott, T. J. The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. Prog. Neurobiol. 90, 385-417 (2010).

SUPPLEMENTARY FIGURE 1



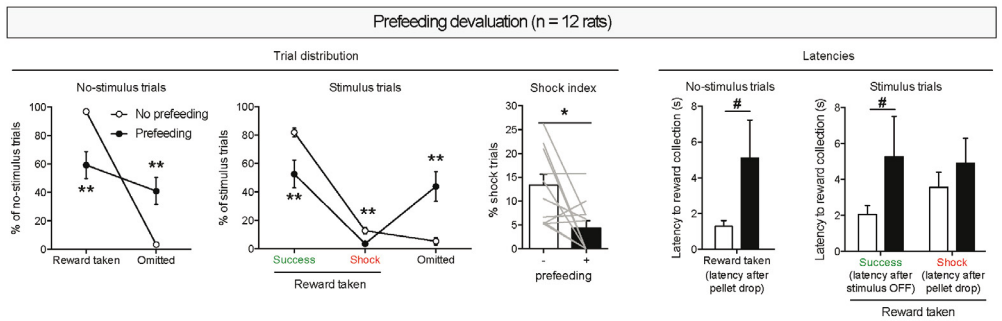
Behavior during stimulus trials over the 12 training sessions in a batch of n = 20 rats (animals from infralimbic and medial orbitofrontal cortex groups). On average, after the 8th training session, animals waited successfully during the majority of stimulus trials. After the 12th session, the experiment commenced. Asterisks denote that this session was the second training session of a day.

SUPPLEMENTARY FIGURE 2



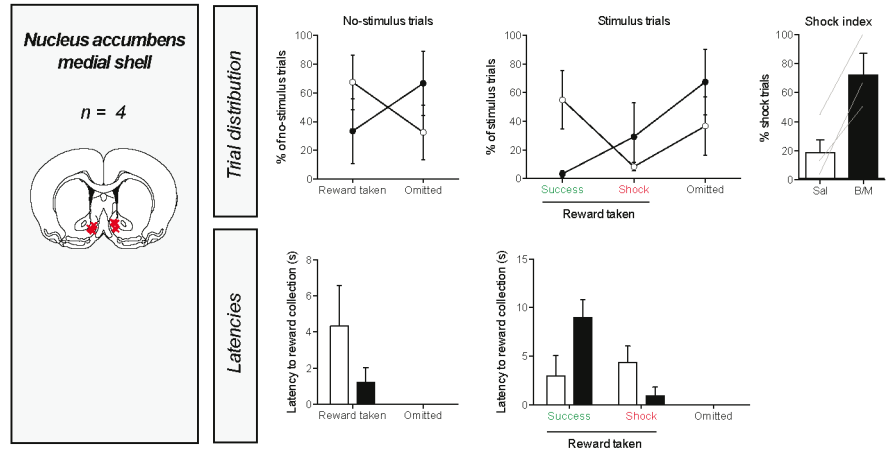
Latencies of pellet retrieval for the different trial types. **** P < 0.0001, *** P < 0.001; see Supplementary statistics table.

SUPPLEMENTARY FIGURE 3



Behavior of the animals after devaluation of reward by selective prefeeding with sucrose pellets matched the expected phenotype of 'loss of motivation'. Gray lines in shock index graph indicate individual animals. # P < 0.1, * P < 0.05, ** P < 0.01 in paired t-test; see Supplementary statistics table.

SUPPLEMENTARY FIGURE 4



Same as figure 4a, but only with animals for which the infusions were in the medial shell subregion of the ventral striatum.

CHAPTER 7

Dopaminergic contributions to behavioral control under threat of punishment in rats

Jeroen P.H. Verharen
Mieneke C.M. Luijendijk
Louk J.M.J. Vanderschuren*
Roger A.H. Adan*

* Equal contribution

Manuscript in preparation

Highlights

- We studied the involvement of the dopamine system in behavioral inhibition by utilizing a recently developed behavioral task for rats
- Midbrain dopamine neurons encode reward prediction error during the task, rather than salience, movement or inhibition of movement
- Neuronal manipulation experiments provide little evidence that midbrain dopamine neurons directly mediate behavioral control

Techniques

CHAPTER 7

Excessive intake of rewards, like food and drugs, often has explicit negative consequences. Thus, choosing not to pursue reward is the result of a cost/benefit decision, execution of which requires behavioral control. Although the dopamine system has been implicated in behavioral control, the neural underpinnings of this process are incompletely understood. Therefore, we studied the involvement of midbrain dopamine neurons and their output regions in behavioral inhibition under threat of punishment, by employing a recently developed 'control over behavior' task for rats. Using *in vivo* fiber photometry, chemogenetics, *c-Fos* immunohistochemistry and behavioral pharmacology, we found little evidence for a direct involvement of midbrain dopamine neurons in inhibitory control over behavior. Rather, the dopamine system seemed to have a role in the motivational component of pursuing reward. Together, our data provide new insights into the mesocorticolimbic mechanisms behind motivated behaviors by showing a modulatory role of dopamine in the expression of cost/benefit decisions.

Introduction

Inhibitory control over behavior is a process that can help to limit the pursuit of reward, and thereby prevent the occurrence of explicit negative consequences that are associated with the excessive intake of reward. In humans, this may for example be the ability to limit the intake of tasty foods in order to prevent obesity, or the ability to refrain from using alcohol and drugs in order not to develop addiction. To study the process of behavioral inhibition in the face of possible punishment, we recently developed a task that studies control over the intake of sucrose pellets in rats[1]. In this task, behavioral control is required during the presentation of an audiovisual threat signal, where an inability to resist the temptation to eat the pellet during this threat signal is punished by a mild electric foot shock. Employing this paradigm, we showed that activity in the ventromedial region of the rat prefrontal cortex (vmPFC) is essential for the exertion of behavioral control, without any effects on task behavior when the animals could take the reward freely, i.e., without the risk of negative consequences. In contrast, the nucleus accumbens (NAc) was important for the motivational aspects of behavior in this task[1].

Dopamine (DA) has been widely implicated in inhibitory control over behavior[2-5]. For example, high trait impulsivity in humans has been associated with low DA release in the striatum and low DA D2 receptor availability[6,7], and monoamine reuptake inhibitors are the cornerstone in the treatment of impulse control disorders like ADHD. Furthermore, functional manipulations of the DA system affect levels of impulsive action[8,9] and impulsive choice[10,11] in rodents, suggesting a causal role of DA neurotransmission in control over behavior. However, the exact mechanism by which DA or its target regions exert control over behavior remains elusive, and it is unknown whether DA neurons are directly engaged during the execution of behavioral control. Importantly, both the vmPFC and the NAc, that play complimentary roles in performance of our behavioral inhibition task[1], receive dense DAergic inputs[12].

Here, we employed a multidisciplinary approach, combining behavioral pharmacology, fiber photometry, chemogenetics and *c-Fos* immunohistochemistry to study the involvement of the mesocorticolimbic DA system in control over behavior in rats. We hypothesized that ventral tegmental area (VTA) DA neurons directly mediate task behavior, by altering DA release in downstream regions. We predicted an important role of mesocortical DA in the exertion of behavioral control and of mesolimbic DA in the motivational aspects of the task, based on the phenotypes observed after pharmacological inactivation of the vmPFC and NAc, respectively[1].

Materials and methods

Animals

A total of 74 male rats with a Long-Evans background, either wild-type Rj:Orl (Janvier labs, France; for *c-Fos* and intracranial infusion experiments) or TH::Cre rats (bred in-house; for photometry and chemogenetics experiments), weighing at least 250 grams at the start of the experiments, were used. Rats were housed in pairs on a 12h/12h reversed day-night cycle (lights off at 8 A.M.). After surgery, animals that received a head implant (photometry and intracranial infusion experiments) were housed individually to prevent damage to the implant. All experimental procedures were conducted in agreement with Dutch laws (Wet op de Dierproeven, 2014) and European guidelines (2016/63/EU) and approved by the Animal Ethics Committee of Utrecht University and the Dutch Central Animal Testing Committee.

Surgeries

Animals were anesthetized by an intramuscular injection of a cocktail of 0.315 mg/kg fentanyl and 10 mg/kg fluanisone (Hypnorm, Janssen Pharmaceutica, Belgium). They were then placed in a stereotaxic apparatus (David Kopf, United States), an incision was made along the midline of the skull and craniotomies were made above the areas of interest:

VTA	AP -5.4 mm	ML ±2.2 mm	DV -8.9 mm from skull under a 10° angle
NAc (core)	AP +1.2 mm	ML ±2.1 mm	DV -6.3 mm from skull under a 5° angle
NAc (shell)	AP +1.2 mm	ML ±2.7 mm	DV -7.0 mm from skull under a 10° angle
vmPFC	AP +3.2 mm	ML ±0.6 mm	DV -3.8 mm from skull

For the NAc and vmPFC, these dorsoventral coordinates reflect the position to which the guide cannulas were lowered; for the VTA, these coordinates reflect the site of viral injection.

For the intracranial infusion experiments, either one 23G guide cannula was used that had a double protrusion, spaced 1.2 mm apart (for the vmPFC; Plastics One, United States), or two 23G guide cannulas with a single protrusion (for the NAc; Plastics One, United States) were used. Guide cannulas were lowered to the desired coordinates, secured with screws, dental glue (C&B Metabond, Parkell Prod Inc., United States) and dental cement, and the skin around the cemented cap was sutured. Dummy cannulas were placed inside the guide cannulas.

For fiber photometry, 1 µl of AAV5-FLEX-hSyn-GCaMP6s or AAV5-hSyn-eYFP (University of Pennsylvania Vector Core; 10^{12} particles/ml), was injected into the right VTA of TH::Cre rats and an optic fiber (diameter 400 µm; Thorlabs, Germany) was lowered to 0.1 mm dorsal of the injected area and secured with screws and dental cement. For chemogenetic experiments, 0.5 µl of AAV5-hSyn-DIO-hM3Gq-mCherry (UCN Vector Core; 2×10^{12} particles/ml) was injected bilaterally into the VTA of TH::Cre rats. Virus was infused at a rate of 0.2 µl/min, and needles were kept in place for an additional 5 minutes after infusion to ensure proper diffusion of the virus into the tissue. For these experiments, measurements were conducted at least 4 weeks later to ensure proper levels of viral expression.

After surgery, all animals received carprofen for pain relief (5 mg/kg, 1x/day, for 3 days, subcutaneously) and saline for rehydration (10 ml once, subcutaneously). Animals were allowed to recover for at least a week before behavioral training started.

Behavioral task

The behavioral task has been extensively described in ref. 1. In brief, we used a task that studies the ability of rats to inhibit their urge to approach a visibly present sucrose pellet during the presentation of an audiovisual threat stimulus. The task comprised 60 trials of 40 seconds each, in which at the start of every trial a sucrose pellet was delivered into a food port (Fig. 1a, left panel). In half of the trials, delivery of this sucrose pellet was not paired with any audiovisual cues, which signaled to the animal that it was safe to consume the pellet directly

without any negative consequences (Fig. 1a, right panel, 'no-stimulus trial'). In the other half of the trials, sucrose pellet delivery co-incided with the onset of an audiovisual (tone+light) cue, which lasted for 12 seconds (Fig. 1a, right panel, 'stimulus trial'). In these trials, the rat had to wait with entering the food port until stimulus termination, thus inhibiting the impulse to consume the sucrose pellet. If the rat managed to do so, it was allowed to take the pellet without further consequences ('success trial'). If the animal was not able to wait and entered the food port during the stimulus, thus losing control over behavior, the stimulus terminated and the animal received a short foot shock punishment (0.3s; 'shock trial'). The intensity of this foot shock was determined for each animal separately during the training phase, but kept constant within the same animal throughout the experiment. Animals typically showed 'attract and repel' behavior with regards to the sucrose pellets (Fig. 1b), likely reflecting behavioral conflict[1]. For the behavioral data, a shock index was computed, which represents the amount of shock trials as a fraction of the amount of shock+success trials, i.e., it is a measure for the amount of shock trials as a function of the total stimulus trials, corrected for the number of omissions.

Experimental procedures

Experimental procedures are described in ref. 1. In brief, behavioral training and testing took place in the dark phase of the reversed day/night cycle. The behavioral task was conducted in operant conditioning chambers (MedPC, Med Associates Inc., United States), equipped with a food port with an infra-red movement detector, flanked by two cue lights, a pellet

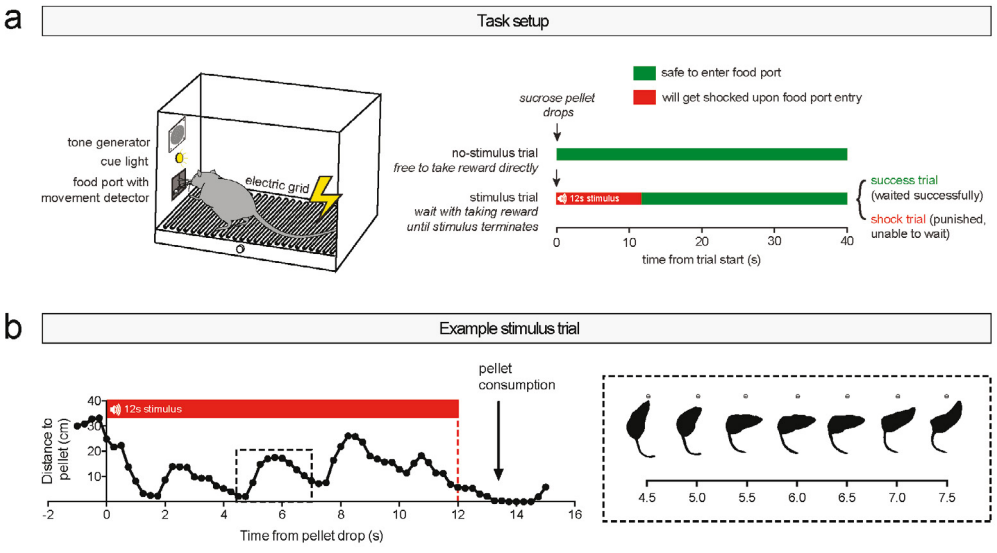


Figure 1: Behavioral control task

a. Behavioral setup. Animals received 60 sucrose pellets at a fixed interval of 40 seconds. Half of the trials were 'no-stimulus' trials, in which the animals could directly retrieve the pellet without negative consequences. The other half of the trials were 'stimulus' trials, in which pellet delivery co-incided with the presentation of an audiovisual threat signal. During this threat signal, food port entry was punished with an electric foot shock.

b. Quantification of behavior in an example stimulus trial, demonstrating 'attract and repel' behavior towards and away from the food port during behavioral inhibition.

Figure modified from ref. 1.

dispenser delivering 45 mg sucrose pellets (SP; 5TULl; TestDiet, United States), a tone generator, a house light and a shock grid floor.

During behavioral training, animals were kept on a food restriction regimen of 4 gram chow per 100 gram body weight, but had *ad libitum* access to tap water in the home cage. Animals were trained for 5-7 days a week, and received one or two training sessions per day. In the first training phase, animals learned to retrieve a sucrose pellet that was delivered into the pellet dispenser at a fixed interval of 40s; this was essentially the final task version but without the stimulus trials. The group of animals progressed to the second and final training phase when all animals retrieved the pellet in at least 55 of the 60 trials. In the final training phase, animals received the regular version of the task, and foot shock intensity was initially set at 0.35 mA. Foot shock intensity was gradually increased with 0.05 or 0.1 mA between sessions when the majority of stimulus trials was punished (punishment too mild), and was decreased if the majority of trials was omitted (punishment too intense). The foot shock intensity was kept constant for an animal when at least 20 out of 30 stimulus trials were 'success' trials. During behavioral testing, animals were food restricted for ~3h prior to the task.

Fiber photometry

Fiber photometry was conducted with a custom-built single wavelength fiber photometry system, as described in ref. 13. In brief, blue 490 nm LED light (Thorlabs, USA) was lock-in amplified (Amplifier SR810; Stanford Research Systems, USA) and delivered through a patch cord (400 µm core diameter; Thorlabs, USA), connected to a stereotaxically placed optic fiber (400 µm diameter; Thorlabs, USA) using a 2.5 mm ceramic ferrule (Thorlabs, USA). Green emission light traveled back through the patch cord, was passed through a dichroic mirror (Semrock, USA) and detected by a photodetector (Newport Corporation, USA). The signal was then passed on to the lock-in amplifier and digitized (Digidata 1550a; Molecular Devices, USA). Next, the raw signal was converted to dF/F values by normalizing each time point F_x to the baseline F_0, which was defined as the average of the 50% middle values of the 30 seconds preceding each time point F_x. We then re-aligned the dF/F traces to the average latencies of pellet retrieval of all 6 animals, so that the different behaviors could be time-locked into one single graph, as has been done in ref. 14. This was accomplished by compressing or stretching the dF/F signal of every trial so it would fit the average latency of pellet retrieval (the average time between pellet drop and retrieval) of the group, using the Matlab command 'resizem'.

Chemogenetics

Animals were injected i.p. with the hM3Dq ligand clozapine-N-oxide (CNO; NIH Drug Supply Program) at a concentration of 0.5 mg/kg dissolved in saline. After injection, animals were placed back into their home cage for 20-30 minutes, before behavioral testing commenced. For the locomotor test, animals were injected with CNO 10 minutes after the start of the experiment (denoted by an arrow in the graph).

Intracranial infusions

For the infusion experiment, injectors were used that protruded 0.6 (NAc) or 1 (vmPFC) mm beyond the termination point of the guide cannulas. One day before the infusions, animals were habituated to the infusion procedure by infusion of 0.3 µl sterile saline. On testing day, animals received an infusion of 0.3 µl saline or 20 µg of cis-(Z)-α-flupenthixol dihydrochloride (Sigma-Aldrich, The Netherlands) dissolved in 0.3 µl saline (counterbalanced between days). The drug was infused with an infusion pump (Harvard Apparatus, United States), set at a rate of 0.5 µl/min. After infusion, injectors were kept in place for an additional 30 seconds (to allow diffusion of the drug into the tissue), and animals were placed back in the home cage for 10-20 minutes before testing began.

c-Fos immunohistochemistry
For the *c-Fos* experiments, 18 animals were trained on the normal version of the task with 60 trials, comprising 30 stimulus trials and 30 no-stimulus trials. During the test session, half the group received 25 stimulus trials ('stimulus group') and the other half received 25 no-stimulus trials ('no-stimulus group'). In an earlier experiment, we omitted the foot shocks from the stimulus trials (to prevent the foot shocks themselves to induce *c-Fos* expression), however, this directly lead to a dramatic reduction in the number of successfully waited trials (i.e., fast extinction of the inhibition response).

90 minutes after termination of the behavioral task, the animals were transcardially perfused with phosphate-buffered saline (PBS) followed by 4% paraformaldehyde (PFA) in PBS. Brains were post-fixed in PFA for 24 hours at 4°C followed by a 30% sucrose solution at 4°C. Brain sections (40 µm) were cut with a cryostat and were stained for *c-Fos* using a 3,3'-Diaminobenzidine (DAB) protocol. First, the sections were blocked for 60 minutes at room temperature using a mixture of 10% normal goat serum and 0.5% Triton-X in PBS, and were then incubated in primary rabbit antibody directed against *c-Fos* (1:1,000; Cell Signaling) in 3% normal goat serum in PBS, overnight at room temperature. The next day, the sections were washed with PBS and incubated with a secondary biotinylated goat antibody directed against rabbit (1:200; Vector labs) for 120 minutes in 3% normal goat serum at room temperature. Sections were then washed in PBS and incubated in Biotin/Avadin (1:1,000; Vectastain) in PBS for 60 minutes. Afterwards, the sections were stained for 3 minutes using liquid DAB (Dako) with 2% nickel ammonium sulphate. After staining, the sections were dehydrated and mounted with a xylene-based mounting medium.

Sections were photographed using a brightfield microscope (at a 5X magnification; AxioImager M2) and *c-Fos* analysis was performed in a semi-automated fashion using an ImageJ (Version 1.51s) routine. In brief, the microscopic images were Fourier-transformed, and a band-pass filter was applied, band-pass filtering structures of approximately the size of *c-Fos*-expressing nuclei (filter was set between 3-6 pixels). Next, peaks in the band-passed image were found using ImageJ's 'Find maxima' function (threshold was set at 145). For each region of interest, the total number of *c-Fos*-expressing cells and the surface area were given, which were used to compute the density of *c-Fos* in that region of interest.

Histological verification
After behavioral experiments, animals were transcardially perfused and brains were sliced according to the protocol described above in paragraph '*c-Fos* immunohistochemistry'. For chemogenetic experiments, VTA sections (50 µm) were cut using a cryostat and stained for hM3Dq and tyrosine hydroxylase (TH) by using free-floating immunohistochemistry. First, sections were blocked for 60' using 3% normal goat serum and 0.3% Triton-X in PBS, and then overnight incubated at 4° using primary antibodies (1:1,000) directed against mCherry (rabbit anti-dsRed; Clontech Living Colors #632496) and TH (mouse anti-TH; Millipore #MAB318) in blocking solution. The next day, sections were washed in PBS and incubated for 120' with secondary antibodies (1:1,000) against rabbit (goat anti-rabbit 568; Abcam #175471) and mouse (goat anti-mouse 488; Abcam #150113). Brain slices were then mounted and coverslipped using FluorSave (Merck Millipore, USA). Images were photographed using an epimicroscope to ensure bilateral expression of the hM3Dq-mCherry. For histological verification of the infusion sites, brain sections were mounted and colored with 5% Giemsa (Sigma-Aldrich, The Netherlands) dissolved in distilled water.

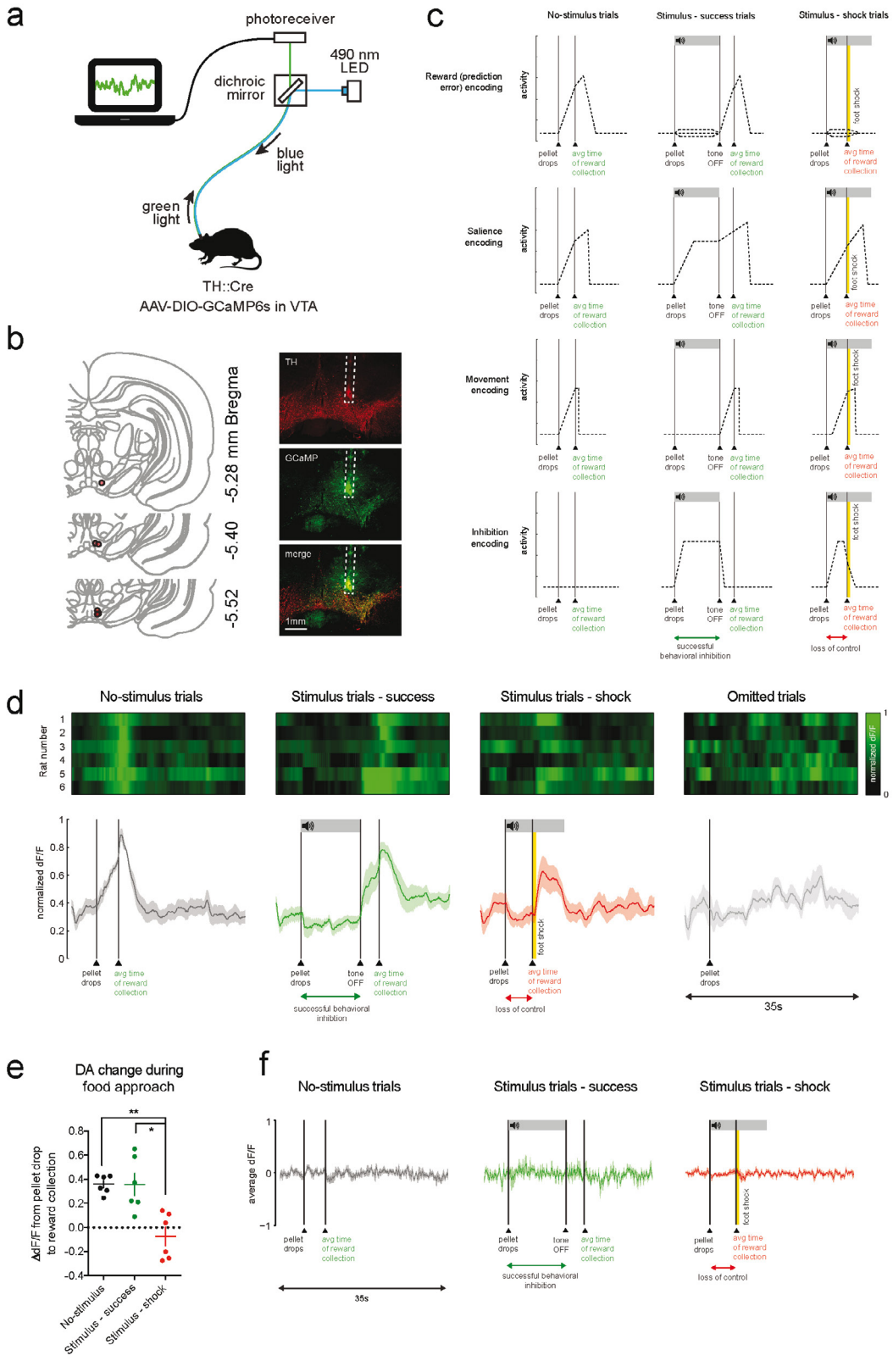
Exclusion criteria
Histological verification of infusion sites and viral expression was performed by an experimenter blind to the experimental results. One animal from the vmPFC infusion group was excluded based on misplacement of the cannulas. One animal from the NAc infusion group was excluded because it lost its head cap. Four animals were excluded from the VTA hM3Dq group because of unilateral expression (2 animals), no expression (1 animal) or hydrocephalus (1 animal). Two animals were excluded from the *c-Fos* experiment; one animal because it was hydrocephalic and one animal because the *c-Fos* staining had not worked (presumably because of an experimental mistake during the staining process). Infusions in the NAc were initially separately targeted at the NAc shell and core, but these groups were eventually combined because the areas were difficult to histologically distinguish, and it was unclear whether the infused volume remained restricted to these NAc subregions.

Code availability
The MedPC script of the task is available at http://www.github.com/jeroenphv.

Statistics
Statistical tests were performed with Prism 6.0 (GraphPad Software Inc., United States). For the dF/F response to food approach of the photometry experiment, a one-way repeated measures of analysis of variance (ANOVA) was used, with stimulus type as a within-subjects repeated measures factor, followed by a Bonferroni post-hoc test. For the locomotor test, a two-way repeated measures ANOVA was used in the time-bin analysis (with time-bin as a within-subjects repeated measures factor and genotype as a between-subjects factor) and an unpaired t-test in the cumulative distance moved analysis. For the data of the behavioral control task, individual paired t-tests were used to compare treatment (CNO or α-flupenthixol) with baseline (saline). For the *c-Fos* experiment, a two-way ANOVA was used with brain area and group as between-subject factors. In all figures, the statistical range is denoted as: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. All test statistics are presented in the Supplementary statistics table.

Results
VTA DA neurons encode reward, but not control over behavior
To study the activity of midbrain DA neurons during successful and unsuccessful control over behavior, we measured population activity of VTA DA neurons by employing *in vivo* fiber photometry[15] in TH::Cre rats (Fig. 2a,b). Based on the different theories of DA function in the brain, we formulated four hypotheses about the expected activity patterns (Fig. 2c). First, DA neurons may encode reward or reward prediction error[16,17], resulting in increased activity when animals can retrieve the pellet without punishment (i.e., after pellet drop in no-stimulus trials or after tone offset in stimulus trials). During stimulus presentation, neurons could either show increased or decreased activity, as both a reward and threat signal are presented, or a combination of the two, which may lead to a net unchanged signal. Second, DA neurons could fire in response to any salient event[18], thus increasing activity in response to pellet delivery, as well as stimulus presentation. Third, we hypothesized that neurons encode movement towards the pellet[19]. Finally, neurons may directly encode inhibition of movement[14,20].

To be able to make a direct comparison between the different animals and trial types, we re-aligned the neuronal population activity to the average response latencies of the animals, by compressing or stretching the dF/F signal (see ref. 14). This analysis revealed a neuronal activation pattern (Fig. 2d) that is reminiscent of the pattern expected based on the reward prediction error hypothesis (Fig. 2c, top panel). During 'No-stimulus' trials, in which the animals were free to take the sucrose pellet directly without negative consequences, we observed a ramping of DA neuron activity from pellet presentation to retrieval, with a decline in activity back to baseline afterwards. Similarly, during 'Stimulus - success' trials, in which animals showed successful control over behavior, we observed this same ramping after tone offset, i.e., when animals were free to take the pellet without negative consequences. No changes in DA neuron activity were observed during successful behavioral control. During 'Stimulus - shock' trials, in which animals retrieved the pellet



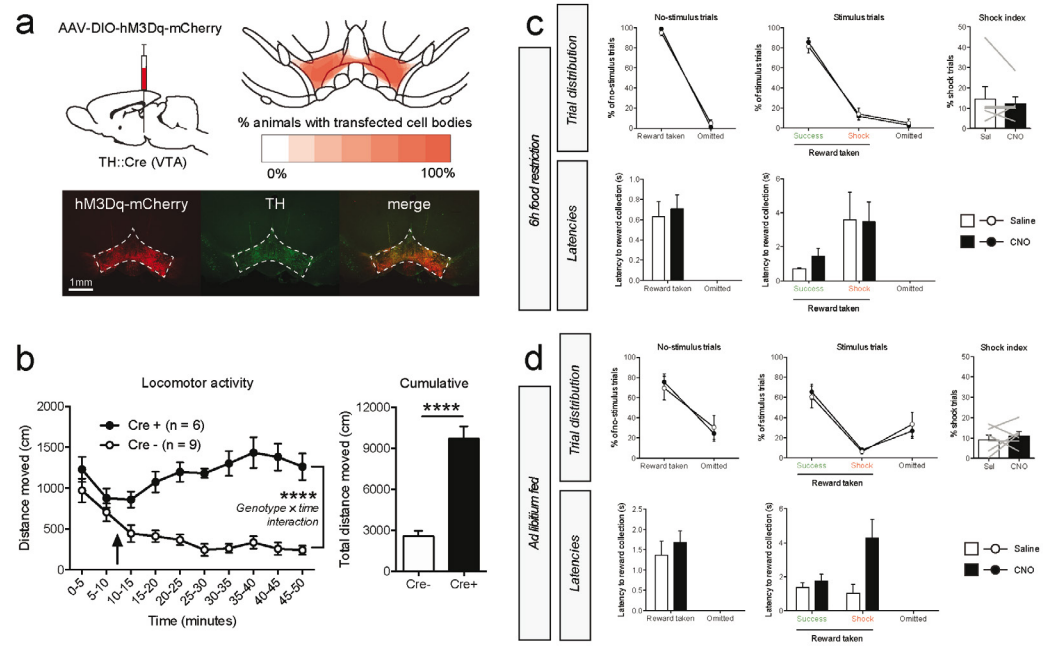


Figure 3: VTA DA neuron activation did not affect task performance. **a.** Experimental procedure and histological verification of hM3Dq-mCherry expression. Bottom and top right images represent coronal slices of the VTA. **b.** Locomotor test. Arrow indicates i.p. CNO injection (0.5 mg/kg). **** P < 0.0001 (see Supplementary statistics table). **c.** Task performance after i.p. CNO injection in hungry animals (n = 6 rats). The shock index is a measure for the relative amount of shock trials compared for the number of omissions, and is computed by 100% * shock trials/(shock + success trials). Gray lines in the shock index graphs indicate individual animals. **d.** Task performance after CNO injection in ad libitum-fed animals (n = 6 rats). Gray lines in the shock index graphs indicate individual animals.

during stimulus presentation and thus received foot shock punishment, we observed a similar response during the inhibition period as during 'Stimulus - success' trials, i.e., no changes in DA neuron activity during stimulus presentation. We observed an increase in DA neuron activity after foot shock delivery, which is something we have observed before[13] and perhaps reflects the salience of the shock. Finally, omitted trials, in which the animals did not retrieve the food pellet during the entire 40s trial period, did not evoke any detectable changes in DA neuron activity. Comparing the changes in dF/F value during approach to the sucrose pellet (Fig. 2e) demonstrated higher dF/F responses to food approach during 'No-stimulus' and 'Stimulus - success' trials as compared to 'Stimulus - shock' trials, suggesting that these DA neurons did not merely respond to movement. Importantly, no changes in fluorescence were observed in animals that were injected with a control fluorophore (Fig. 2f). In sum, these data suggest that VTA DA neurons encode reward or reward prediction error, but not (successful or unsuccessful) behavioral control or movement.

← **Figure 2:** VTA DA neurons encoded reward or reward prediction error during the task. **a.** Fiber photometry setup. **b.** Histological verification. Red circles indicate fiber placement of individual animals. **c.** Expected activity patterns. **d.** VTA DA neuron activity during the different trial types (n = 6 rats). Bottom graph shows mean ± standard error of the mean of the six animals. **e.** Quantification of dF/F signal during food approach. ** P = 0.0098, * P = 0.0105 in post-hoc t-tests (see Supplementary statistics table). **f.** dF/F of animals injected with control fluorophore eYFP (n = 4 rats).

VTA DA neuron activation does not affect task performance

To assess whether hyperactivity of these same VTA DA neurons hampers the exertion of behavioral control, we injected TH::Cre rats with a viral vector expressing the excitatory chemogenetic receptor hM3Dq fused to mCherry-fluorescent protein bilaterally into the VTA (Fig. 3a). To confirm functional activation of these neurons, we assessed locomotor activity[21] after injection of the hM3Dq ligand clozapine-N-oxide (CNO) and observed an increase in the distance traveled in animals that were TH::Cre positive as compared to TH::Cre negative animals (Fig. 3b).

Contrary to expectations, we observed no effects of chemogenetic VTA DA neuron activation on task performance in food-restricted animals (Fig. 3c). Given that food restriction increases baseline firing of DA neurons[22,23], we speculated that firing in these neurons could already have been high before CNO injection, and this may have therefore masked an effect on task behavior. Therefore, we repeated the experiment in *ad libitum*-fed animals, but we again observed no effects of neuronal activation on task performance (Fig. 3d). These findings indicate that increasing the activity of VTA DA neurons does not impair the ability of animals to exert inhibitory control over behavior.

Stimulus trials engage the NAc, but not vmPFC

To explore whether the two major VTA DA output regions, the NAc and vmPFC, are recruited during stimulus trials, we tested a group of animals in a task version that comprised exclusively no-stimulus trials and a different group of animals in a task version that comprised exclusively stimulus trials, and stained the brain sections for the immediate early gene *c-Fos*, as a proxy for neuronal activity[24] (Fig. 4a). We then performed semi-automated cell counting on two coronal slices that included the vmPFC and NAc, and compared the cumulative density of *c-Fos* levels in these brain regions.

A two-way analysis of variance (ANOVA) on the *c-Fos* density in the three major regions of the NAc revealed a significant main effect of group, but no group × brain region interaction effect, indicating that *c-Fos* expression was increased across the entire NAc after stimulus trials (Fig. 4b). In contrast, no effects of group or a group × brain region interaction effect were found for *c-Fos* density in the vmPFC (Fig. 4c). Together, these findings suggest that the NAc, but not the vmPFC, is recruited during stimulus trials.

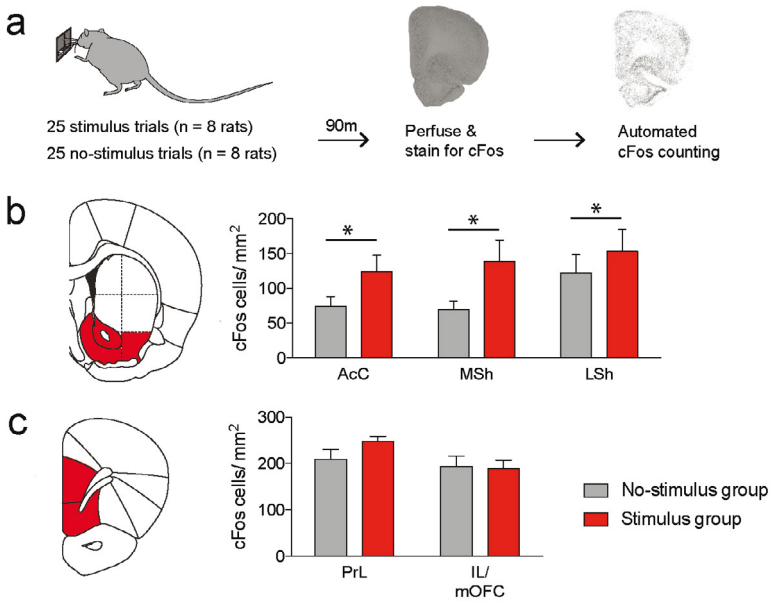


Figure 4: Animals that received stimulus trials showed enhanced c-Fos expression in the NAc, but not vmPFC, as compared to animals that received no-stimulus trials.
a. Experimental procedure.
b. c-Fos density was enriched across the entire ventral striatum after stimulus trials (* P = 0.0153, main effect of group in ANOVA; see Supplementary statistics table). Abbreviations: AcC, nucleus accumbens core; MSh, medial shell of the nucleus accumbens; LSh, lateral shell of the nucleus accumbens.
c. Stimulus trials did not evoke changes in c-Fos density in the vmPFC (See Supplementary statistics table). Abbreviations: PrL, prelimbic cortex; IL, infralimbic cortex; mOFC, medial orbitofrontal cortex.

Blockade of DA receptors in the NAc and vmPFC affects task performance

To investigate the importance of DAergic neurotransmission in VTA target regions for performance in the task, we tested the effects of infusion of the DA receptor antagonist

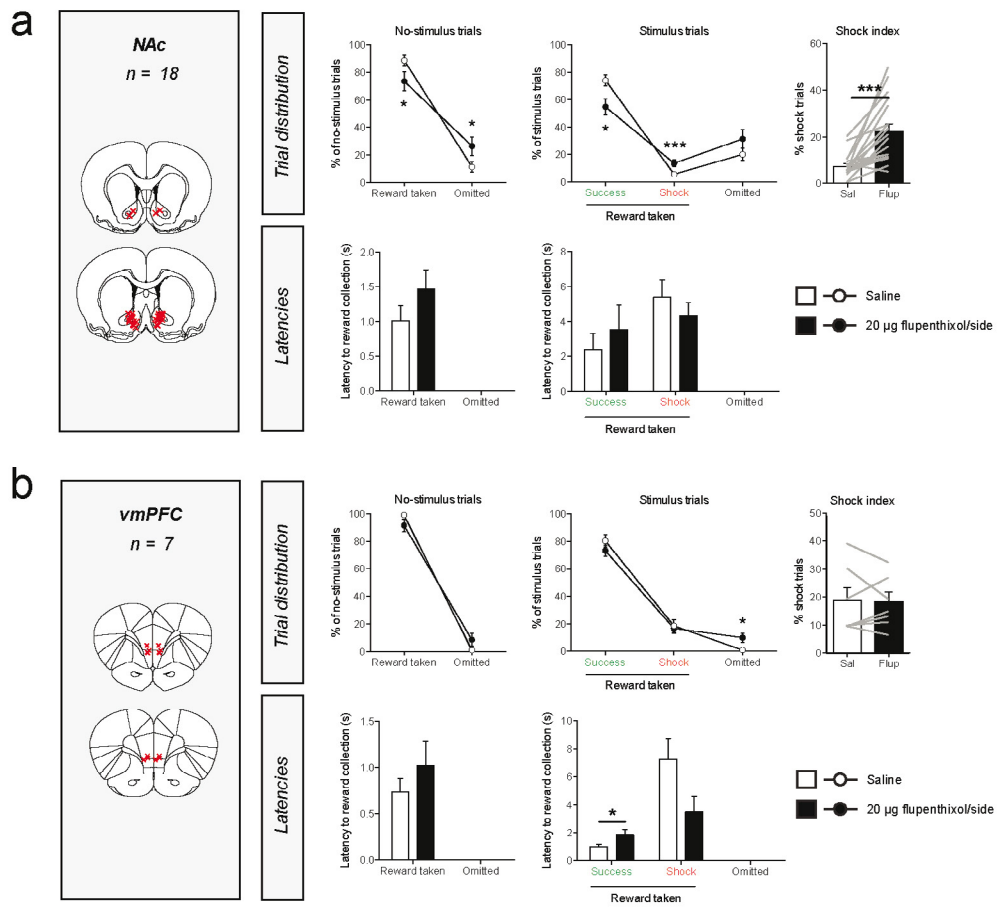


Figure 5: DA receptor blockade by intra-NAc **(a)** or intra-vmPFC **(b)** infusion of α-flupenthixol had differential effects on task performance. Red crosses in the coronal brain sections represent the infusion sites in each experiment. Gray lines in the shock index graphs indicate individual animals. *** P < 0.001, * P < 0.05 in paired t-test (see Supplementary statistics table).

α-flupenthixol into the NAc and vmPFC (Fig. 5). During no-stimulus trials, we observed a significant increase in the number of omissions after α-flupenthixol infusion into the NAc (Fig. 5a). In stimulus trials, we observed a significant decrease in the number of success trials and a significant increase in the number of shock trials, but no significant effect on the number of omissions. Hence, the shock index was significantly increased after α-flupenthixol infusion. No effects were observed on the latency of pellet retrieval in either trials. Infusion of α-flupenthixol into the vmPFC had no effects on behavior during no-stimulus trials (Fig. 5b). In stimulus trials, it resulted in a significant increase in the number of omissions, but no significant changes in the number of success or shock trials. We further observed a significant increase in the latency of pellet retrieval in success trials, but not in shock trials.

Discussion

In this study, we have utilized a recently developed task to assess the contribution of the mesocorticolimbic DA system to inhibitory control over behavior in rats. By combining this task with *in vivo* fiber photometry, chemogenetics, *c-Fos* immunohistochemistry and behavioral pharmacology, we have provided novel insights into the role of DA in behavioral control and other aspects of task performance. First, we have visualized neuronal dynamics *in vivo* during the moment that animals demonstrated successful and unsuccessful control over behavior by employing fiber photometry in TH::Cre rats. This experiment indicated that VTA DA neuron activity most closely represented a pattern of reward prediction error coding, rather than movement or inhibition of movement. When the animals successfully inhibited the urge to consume sucrose, VTA DA neuron transients remained unchanged, suggesting that the presence of the audiovisual threat cue suppressed the occurrence of the positive reward signal evoked by delivery of the sugar pellet. The photometry fibers were for the most part placed in the medial aspect of the paranigral nucleus of the VTA. As a result, our measurements were mostly from DA neurons projecting to the NAc core and PFC[25-27]. We can therefore not state that the observed neuronal dynamics are representative of the entire population of midbrain DA neurons, especially given the fact that fiber photometry measures aggregate fluorescence, i.e., this technique does not allow for the detection of functional heterogeneity in the recorded neuronal population. That said, chemogenetic activation of VTA DA neurons did not hamper inhibitory control in the task, which supports the notion that these neurons do not directly govern behavioral control. Consistently, we have previously found that chemogenetic activation of VTA DA neurons in TH::Cre rats did not increase motor impulsivity in the 5-choice serial reaction time task[28]. Furthermore, we have shown that hyperactivation of VTA DA neurons impaired the ability of rats to behaviorally adapt to negative, but not positive, reward prediction errors[13]. This may explain why we did not observe any effects on task behavior after CNO injection, as the photometry data indicated the presence of exclusively positive DA neuron transients during the task.

The results of the *c-Fos* expression experiment showed increased neuronal activation in the NAc, but not the vmPFC, in animals that received exclusively stimulus trials compared to animals that received exclusively no-stimulus trials in the task. This suggests that the NAc is actively engaged during the execution of stimulus trials, although it is not possible to determine whether this is the direct result of the NAc mediating behavioral control, or that it is due to other aspects of stimulus trials, like the continuous threat of foot shock[29]. Furthermore, it is interesting to note that we did not observe any significant changes in *c-Fos* expression in the vmPFC, given that we have previously demonstrated impairments in behavioral control after inactivation of this region[1]. That said, the exertion of behavioral control by the vmPFC does not necessarily have to be the result of a simple increase in the region's activity, but may as well be due to more subtle changes in activity, such as alterations in the canonical computations within the vmPFC or of changed activity in a small subpopulation of vmPFC neurons, which would logically not result in increased *c-Fos* expression across the entire region.

DA receptor blockade in the NAc significantly decreased the number of success trials and increased the absolute number of shock trials. Although this effect was numerically more modest than the phenotype observed after pharmacological inactivation of the vmPFC[1], it does suggest decreased inhibitory control. However, in no-stimulus trials, which can be seen as a control to detect any general impairments in behavior, an increase in omissions was observed. This suggests that additional cognitive processes that are necessary for correct task execution were compromised, such as motivation or attention. Therefore, the effects of NAc DA receptor antagonism on behavioral control should be interpreted with caution. Indeed, most studies report no effect of DA receptor blockade in the NAc on classical measures of impulsive action and impulsive choice[3,30], suggesting that the observed pattern of effects after DA receptor antagonist infusion was not primarily driven by changes in behavioral control, but rather by the disruption of other cognitive processes. For example, it could be the case the DA released during reward prediction, as we have shown with our photometry experiment, cannot be detected by the NAc, which may lead to alterations in motivation or impairments in the detection of pellet delivery and stimulus presentation.

In contrast to the NAc, pharmacological blockade of DA receptors in the vmPFC did alter behavior during stimulus trials, while not affecting behavior during no-stimulus trials. Interestingly, this did not seem to be related to behavioral control, but rather by a decreased motivation for reward in stimulus trials. As such, we observed an increase in the number of omissions and an increased latency of pellet retrieval in success trials. This phenotype is different than the one induced by pharmacological inactivation of the vmPFC, which was characterized by impairments in inhibitory control[1], indicating that the role of the vmPFC in inhibitory control is not governed by DAergic neurotransmission. Instead, the isolated motivational effects of α-flupenthixol on stimulus, but not no-stimulus trials, perhaps suggests increased task-related anxiety with regards to retrieving the pellet after stimulus presentation. This could in turn be the result of the overexpectation of negative consequences, for example because of disturbances in weighing the costs and benefits of different courses of actions, a function that has been attributed to mesocortical DA[31].

In sum, we have used a multidisciplinary approach to test the hypothesis that mesocorticolimbic DA is involved in the exertion of inhibitory behavioral control over food intake. We found little evidence in support of this hypothesis, as we did not observe changes in VTA DA neuron activity during successful and unsuccessful behavioral control. Furthermore, chemogenetic DA neuron activation did not affect task performance. We did find that stimulus trials lead to increased *c-Fos* expression in the NAc, and DA receptor blockade within the NAc resulted in an increased amount of shock trials, but this may not necessarily have been the result of a direct impairment in the ability to exert behavioral control. Furthermore, DA receptor blockade in the vmPFC did not change measures of inhibitory control, even though we have previously shown that activity in this area is essential for this behavior[1]. Our findings contribute to the understanding of the role of DA in motivated behaviors. That is, DA is not likely to be a direct mediator of the type of behavioral inhibition that is assessed in our task.

References

1. Chapter 6 of this thesis.
2. Cools, R. Role of dopamine in the motivational and cognitive control of behavior. Neuroscientist 14, 381-395 (2008).
3. Pattij, T. & Vanderschuren, L. J. The neuropharmacology of impulsive behaviour. Trends Pharmacol Sci 29, 192-199 (2008).
4. Dalley, J. W. & Roiser, J. P. Dopamine, serotonin and impulsivity. Neuroscience 215, 42-58 (2012).
5. Nutt, D. J., Lingford-Hughes, A., Erritzoe, D. & Stokes, P. R. The dopamine theory of

addiction: 40 years of highs and lows. Nature reviews. Neuroscience 16, 305-312 (2015).
6. Buckholtz, J. W. et al. Dopaminergic Network Differences in Human Impulsivity. Science 329, 532-532 (2010).
7. Trifilieff, P. & Martinez, D. Imaging addiction: D2 receptors and dopamine signaling in the striatum as biomarkers for impulsivity. Neuropharmacology 76, 498-509 (2014).
8. van Gaalen, M. M., Brueggeman, R. J., Bronius, P. F., Schoffelmeer, A. N. & Vanderschuren, L. J. Behavioral disinhibition requires dopamine receptor activation. Psychopharmacology 187, 73-85 (2006b).
9. Pattij, T., Janssen, M. C. W., Vanderschuren, L. J. M. J., Schoffelmeer, A. N. M. & Van Gaalen, M. M. Involvement of dopamine D 1 and D 2 receptors in the nucleus accumbens core and shell in inhibitory response control. Psychopharmacology 191, 587-598 (2007).
10. Wade, T. R., de Wit, H. & Richards, J. B. Effects of dopaminergic drugs on delayed reward as a measure of impulsive behavior in rats. Psychopharmacology 150 (2000).
11. van Gaalen, M. M., van Koten, R., Schoffelmeer, A. N. & Vanderschuren, L. J. Critical involvement of dopaminergic neurotransmission in impulsive decision making. Biol. Psychiatry 60, 66-73 (2006a).
12. Bjorklund, A. & Dunnett, S. B. Dopamine neuron systems in the brain: an update. Trends Neurosci 30, 194-202 (2007).
13. Verharen, J. P. H. et al. A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. Nature communications 9 (2018).
14. Syed, E. C. et al. Action initiation shapes mesolimbic dopamine encoding of future rewards. Nature neuroscience (2015).
15. Gunaydin, L. A. et al. Natural neural projection dynamics underlying social behavior. Cell 157, 1535-1551 (2014).
16. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. Science 275 (1997).
17. Schultz, W. Dopamine reward prediction-error signalling: a two-component response. Nature reviews. Neuroscience 17, 183-195 (2016).
18. Berridge, K. C. The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology 191, 391-431 (2007).
19. Howe, M. W. & Dombeck, D. A. Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature 535, 505-510 (2016).
20. Mazzoni, P., Hristova, A. & Krakauer, J. W. Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. Journal of neuroscience 27, 7105-7116 (2007).
21. Boekhoudt, L. et al. Chemogenetic activation of dopamine neurons in the ventral tegmental area, but not substantia nigra, induces hyperactivity in rats. European Neuropsychopharmacology 26, 1784-1793 (2016).
22. Hommel, J. D. et al. Leptin receptor signaling in midbrain dopamine neurons regulates feeding. Neuron 51, 801-810 (2006).
23. Branch, S. Y. et al. Food restriction increases glutamate receptor-mediated burst firing of dopamine neurons. Journal of Neuroscience 33, 13861-13872 (2013).
24. Bullitt, E. Expression of c fos like protein as a marker for neuronal activity following noxious stimulation in the rat. Journal of Comparative Neurology 296, 517-530 (1990).
25. Lammel, S. et al. Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. Neuron 57, 760-773 (2008).
26. Lammel, S., Lim, B. K. & Malenka, R. C. Reward and aversion in a heterogeneous midbrain dopamine system. Neuropharmacology 76, 351-359 (2014).
27. Morales, M. & Margolis, E. B. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. Nature reviews. Neuroscience 18, 73-85 (2017).
28. Boekhoudt, L. et al. Chemogenetic Activation of Midbrain Dopamine Neurons Affects Attention, but not Impulsivity, in the Five-Choice Serial Reaction Time Task in Rats. Neuropsychopharmacology 42, 1315 (2016).
29. Beck, C. & Fibiger, H. Conditioned fear-induced changes in behavior and in the expression of the immediate early gene c-fos: with and without diazepam pretreatment. Journal of Neuroscience 15, 709-720 (1995).
30. Pezze, M., Dalley, J. W. & Robbins, T. W. Differential Roles of Dopamine D1 and D2 Receptors in the Nucleus Accumbens in Attentional Performance on the Five-Choice Serial Reaction Time Task. Neuropsychopharmacology 32, 273 (2006).
31. Floresco, S. B. Prefrontal dopamine and behavioral flexibility: shifting from an "inverted-U" toward a family of functions. Front Neurosci 7, 62 (2013).

CHAPTER 8

Limbic control over the homeostatic need for sodium

Jeroen P.H. Verharen*
Theresia J.M. Roelofs*
Shanice Henry
Mieneke Luijendijk
Rick M. Dijkhuizen
Louk J.M.J. Vanderschuren
Roger A.H. Adan

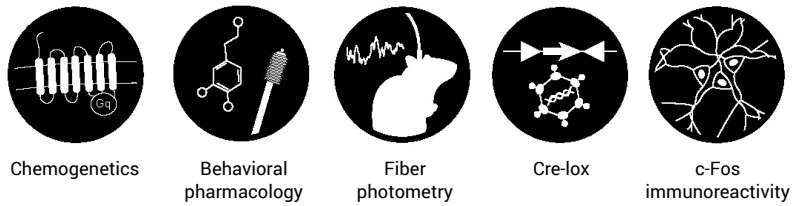
* Equal contribution

Manuscript under review

Highlights

- Salty foods can be aversive or appetitive, depending on the homeostatic state of the body
- Dopamine neurons in the ventral tegmental area of the rat encode this switch in salt appreciation
- The NAc mediates the motivational, but not hedonic, component of sodium appetite, but this is not dopamine dependent

Techniques



CHAPTER 8

The homeostatic need for sodium is one of the strongest motivational drives known in animals. Although the brain regions involved in the sensory detection of sodium levels have been relatively well mapped, data about the neural basis of the motivational properties of salt appetite, including a role for midbrain dopamine cells, have been inconclusive. Here, we employed a combination of fiber photometry, behavioral pharmacology and c-Fos immunohistochemistry to study the involvement of the mesocorticolimbic dopamine system in salt appetite in rats. We observed that sodium deficiency affected the responses of dopaminergic midbrain neurons to salt tasting, suggesting that these neurons encode appetitive properties of sodium. We further observed a significant reduction in the consumption of salt after pharmacological inactivation of the nucleus accumbens (but not the medial prefrontal cortex), and microstructure analysis of licking behavior suggested that this was due to decreased motivation for, but not appreciation of salt. However, this was not dependent on dopaminergic neurotransmission in that area, as infusion of a dopamine receptor antagonist into the nucleus accumbens did not alter salt appetite. We conclude that the nucleus accumbens, but not medial prefrontal cortex, is important for the behavioral expression of salt appetite by mediating its motivational component, but that the switch in salt appreciation after sodium depletion, although detected by midbrain dopamine neurons, must arise from other areas.

Introduction

In order to obtain all nutrients necessary for survival, organisms need to make adaptive food choices based on their homeostatic needs. For example, when an organism's body senses a shortage of a certain nutrient, it may, consciously or not, choose foods that will replenish this need. Of all the nutrients, a deficiency in sodium is one of the strongest homeostatic drives known in animals, evoking intense cravings for salty foods after salt deprivation, which has been consistently reported in a wide range of species[1,2]. Although sodium is abundant in modern Western diets, it is relatively scarce in natural resources, which has perhaps contributed to the development of this homeostatic drive.

A remarkable observation that illustrates this innate drive for sodium is that rats normally experience a hypertonic sodium solution as aversive, but that this solution is experienced as positive and consumed in high amounts when rats are low on sodium, a phenomenon known as salt appetite[1-4]. Such a switch in the experience of a flavor from aversive to appetitive, driven by a homeostatic need, is a prime example of how adaptive the interaction between sensory and reward systems can be in order to maintain homeostasis and ensure survival. Elucidating the mechanisms that underlie this switch may therefore provide interesting insights into the flexibility of brain circuits that mediate reward.

A variety of brain areas has been shown to be involved in salt appetite. Not surprisingly, this includes brain structures involved in the sensory processing of taste, such as the parabrachial nucleus[5] and the nucleus of the solitary tract[6,7]. Other brain areas implicated in salt appetite are the lateral and paraventricular nucleus of the hypothalamus, the preoptic area, the subfornical organ, the central amygdala and the bed nucleus of the stria terminalis (for a review see ref. 8). Given its role in processing rewarding and aversive stimuli[9,10], a logical candidate for the mediation of salt appetite is the mesocorticolimbic dopamine (DA) system, consisting of DA neurons in the ventral tegmental area (VTA) projecting to the nucleus accumbens (NAc) and medial prefrontal cortex (mPFC). However, data about the involvement of this circuit in salt appetite has been inconclusive. On the one hand, a total ablation of the VTA[11] or DA terminals in the entire brain[12], as well as the infusion of DA receptor agonists or antagonists in the nucleus accumbens[13] does not affect salt appetite, suggesting that motivation for salt bypasses the mesoaccumbens DA pathway. On the other hand, it has been observed that infusion of a delta-opioid receptor antagonist into the VTA decreases salt appetite[13], and that a sodium-depleted state is associated with decreased DA transporter activity[14] and altered spine morphology[15] in the nucleus accumbens. A recent study demonstrated, using fast-scan cyclic voltammetry, that tasting a sodium solution evoked phasic dopamine release in the rat nucleus accumbens shell after sodium deprivation, but not under normal conditions[16]. Furthermore, this study showed that hindbrain neurons projecting to the VTA displayed increased c-Fos expression after salt deprivation. Another recent study showed that optogenetic or chemogenetic activation of VTA DA neurons in mice reduced intake of a high-concentration (but not low-concentration) salt jelly, while chemogenetic inhibition of these same neurons had no effect on salt intake[17].

In this study, we attempted to contribute to the understanding of the involvement of the mesocorticolimbic DA system in salt appetite. Towards this aim, we combined fiber photometry, behavioral pharmacology and c-Fos immunohistochemistry to study *in vivo* VTA DA neuron dynamics during sodium deficiency, and the importance of the NAc and mPFC, the two major output regions of these neurons, for salt appetite. By employing a microstructural analysis of licking behavior, we tried to discern effects of manipulations of the mesocorticolimbic system on the motivation for versus the appreciation of salt. We hypothesized that VTA DA neurons may respond differently to salty solutions during a normal versus sodium-depleted state, and that these changes in DA cell responsiveness are important for the expression of behaviors associated with salt appetite.

Results

No changes in c-Fos expression in DA nuclei after sodium deprivation

In an attempt to substantiate the findings of ref. 17, that showed that sodium deprivation did not change baseline activity of VTA DA neurons, we analyzed c-Fos immunoreactivity in a coronal slice of the midbrain that included the VTA and substantia nigra (Fig. 1a). Based on a typical brain slice, we created a template on which we overlayed all the other slices in order to perform whole-slice automated cell counting. A visual sliding-window analysis revealed fairly similar levels of c-Fos expression between animals in a sodium-depleted state (induced by treatment with the diuretic furosemide; see Materials and methods) versus a control state around these nuclei (Fig. 1b). Indeed, region-of-interest analysis showed no significant differences in the number of c-Fos positive cells in the VTA (Fig. 1c), the substantia nigra pars compacta (SNc; Fig. 1d), or the substantia nigra pars reticulata (SNr; Fig. 1e). Together, these data support the finding that baseline activity of neurons in midbrain DA nuclei was not altered by sodium deprivation.

Dopamine neurons encode a switch in sodium appreciation

To study how VTA DA neurons respond to the taste of salt during normal and low levels of sodium in the body, we injected a viral vector carrying Cre-dependent GCaMP6s into the VTA of TH::Cre rats and measured VTA DA neuron dynamics using fiber photometry during a Pavlovian conditioning task. In this task (Fig. 2a), rats learned that a 5-second auditory tone preceded the delivery of a nutritional solution, which was usually a tasty sucrose solution (3 out of 4 trials), but sometimes a NaCl solution (1 out of 4 trials). We tested the responses of the animals to these solutions on two occasions: once in a sodium-deficient state, 24h after injection with furosemide, and once under baseline conditions (Fig. 2b).

In the control state, animals vigorously licked for sucrose, but refrained from

licking when a sodium solution was delivered, in line with our expectations that high-sodium concentrations are aversive to rats. Accordingly, VTA DA neuron population activity increased during the consumption of sucrose, while the delivery of salt resulted in sub-baseline levels of DA neuron activity. This is illustrated in both an example animal (Fig. 2c, left panel) as well as on a group level (Fig. 2d-f, black curves).

In the sodium-depleted state, animals significantly increased licking for salt, while the number of licks for sucrose remained the same as in the control state (Fig. 2f). In line with this appreciation of salt, we observed increased levels of VTA DA neuron activity in response to salt delivery (Fig. 2e), with responses that were even higher than those after sucrose delivery (compare Fig. 2d and e, blue curves). Although numerically more modest than the changed DA neuron responsiveness to salt, we observed a lower DA neuron activation to sucrose during a salt-depleted state compared to the control state (Fig. 2d). Importantly, we

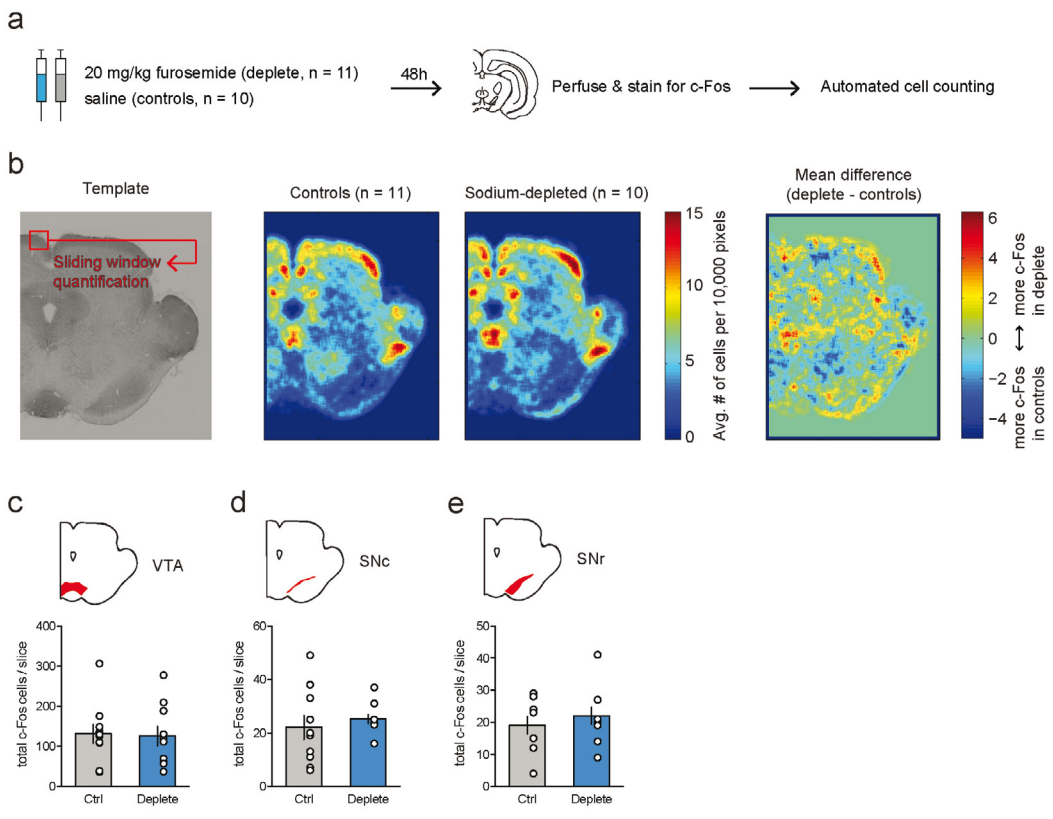


Figure 1 c-Fos analysis of midbrain sections after sodium deprivation. **a)** Experimental design. **b)** From left to right: a coronal section of the midbrain that included the VTA and substantia nigra was used to create a template on which the other midbrain sections were overlayed in order to perform whole-slice automated cell counting – average c-Fos density in control animals (n=11) – average c-Fos density in sodium-depleted animals (n=10) – mean difference in c-Fos expression between controls and depleted animals indicating similar levels of c-Fos expression. **c-e)** Region-of-interest analysis showed no significant differences in the number of c-Fos positive cells in the VTA (**c**; t18 = 0.15, p = 0.88), the substantia nigra pars compacta (SNc; **d**; t18 = 0.64, P = 0.53), or the substantia nigra pars reticulata (SNr; **e**; t18 = 0.75, P = 0.46).

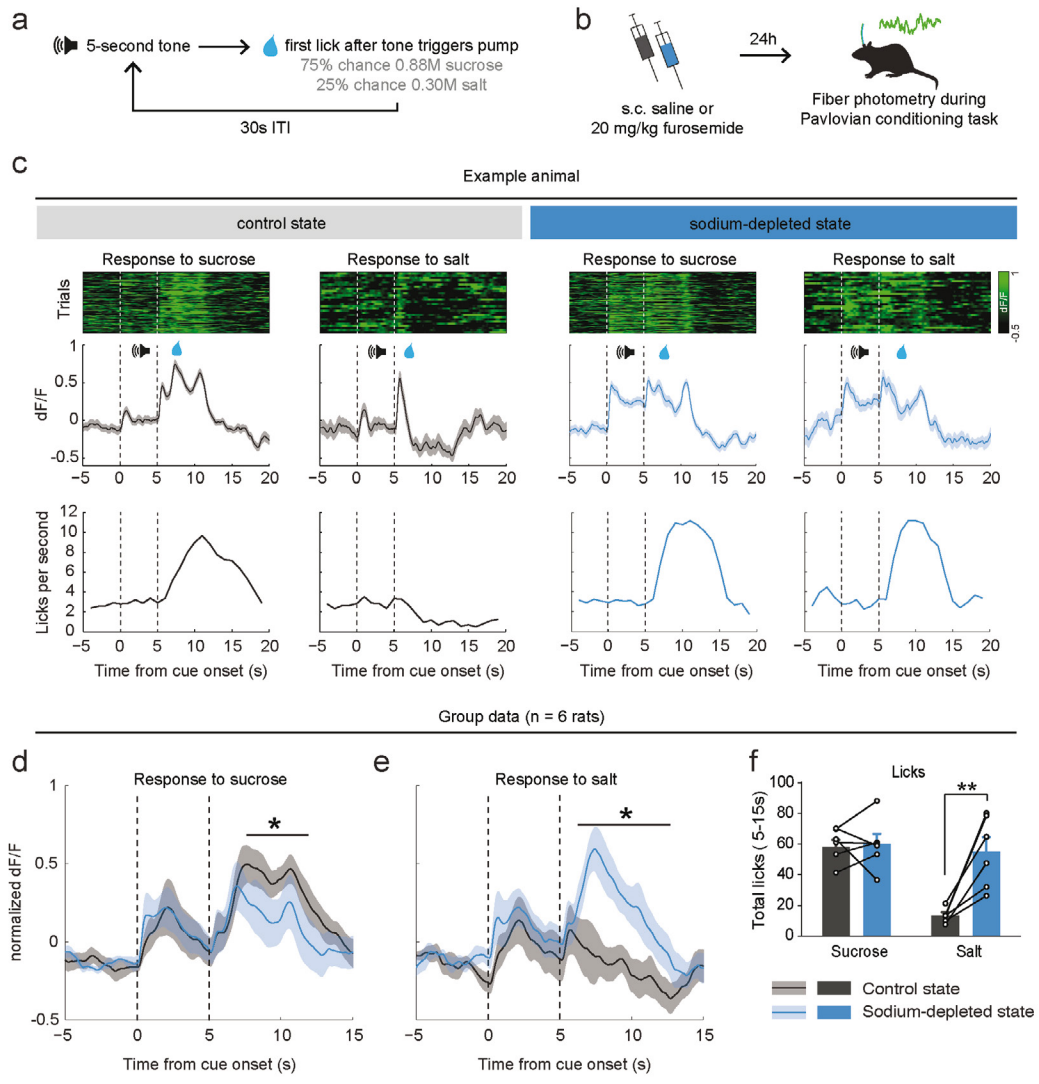


Figure 2 In vivo fiber photometry of VTA DA neurons during sodium depletion. **a)** The Pavlovian conditioning task that was used for in vivo fiber photometry consisted of a five-second auditory tone followed by delivery of a nutritional solution, being either a sucrose solution (in 75% of trials) or a NaCl solution (in 25% of trials). A 30-second inter trial interval (ITI) separated the trial from the next auditory tone. **b)** Animals were tested twice; once after a subcutaneous (s.c.) saline injection, i.e. a control state, and once after a furosemide injection, i.e. a sodium-depleted state. **c)** Population responses of VTA DA neurons of an example animal. Shown are the control state (left) and sodium-depleted state (right). Reward was delivered for 5s after the first lick after cue offset (5s). **d)** Salt depletion decreased VTA DA neuron responses to sucrose. The two lines represent mean responses of all animals in the control (black) and the sodium-depleted (blue) state. A significant main effect of treatment was found (2-way repeated measures ANOVA, main effect of treatment, $F_{1,5} = 2.110$, p = 0.2060; treatment × time interaction effect, $F_{1999,9995} = 0.1.788$, P < 0.0001; * post-hoc test significant between 7.41 - 12.00 s post-stimulus). **e)** Mean responses of all animals to salt indicated that salt depletion increased VTA DA neuron response to salt (2-way repeated measures ANOVA, main effect of treatment, $F_{1,5} = 26.45$, P = 0.0036; treatment × time interaction effect, $F_{1999,9995} = 2.400$, P < 0.0001; * post-hoc test significant between 6.32 - 12.72 s post-stimulus). **f)** Number of licks in the first 10 seconds after sucrose or salt delivery for animals in the control (black) or sodium-depleted (blue) state. Salt depletion increased the number of licks for salt, but not for sucrose (2-way repeated measures ANOVA, main effect of tastant, $F_{1,5} = 21.54$, P = 0.0056; main effect of treatment, $F_{1,5} = 10.74$, P = 0.0220; treatment × tastant interaction effect, $F_{1,5} = 19.93$, P = 0.0066; post-hoc Sidak's test, control vs depleted state: sucrose $t_5 = 0.017$, P = 0.9998; salt $t_5 = 6.297$, ** P = 0.0030).

observed no changes in fluorescent activity in animals that were injected with an activity-independent control fluorophore (Supplementary Fig. 1), indicating that the observed fluorescent signals were driven by neuronal activity. Furthermore, neuronal activity was not driven by licking per se, since during the anticipatory conditioned stimulus, we observed an increase in calcium activity, but not in the number of licks (Fig. 2c).

In sum, we show that a salty solution under normal conditions is considered aversive by rats, as shown by the termination of licking behavior and sub-baseline levels of VTA DA neuron activity, but that this same solution is considered appetitive in a sodium-depleted state, accompanied by vigorous licking for salt and large peaks in DA neuron activity.

Inactivation of NAc, but not mPFC, diminished drinking behavior without affecting salt appetite

To investigate the behavioral structure of salt appetite, we assessed intake of a 0.45M NaCl solution, as well as intake of demineralized water, by using mechanical lickometers present in the animals' home cages, which measured the numbers of licks per 12s bins. To gain insight into the appetitive components of sodium appetite, we performed a microstructure analysis of licking behavior, calculating the number of licking bouts that animals made, as well as the size of each of these bouts (Fig. 3a). As expected, animals that were brought in a sodium-depleted state consistently consumed more of the sodium solution, which was driven by an increase in the frequency of licking bouts as well as the size of a licking bout (see Fig. 3 and 5). Note that these animals had *ad libitum* access to demineralized water, but had no access to salt in the 24h prior to the measurements.

To study the role of the two main VTA DA neuron output regions, the mPFC and NAc, in the regulation of salt appetite, we pharmacologically inactivated the mPFC and the NAc using micro-infusions of a mixture of the GABA receptor agonists baclofen and muscimol (B/M). Rats were brought in a sodium-depleted or a control state for 24h, after which they received infusions with either B/M or saline. Subsequently, animals received a 0.45M NaCl solution, in addition to a bottle of demineralized water that was already present in the cage.

We first assessed salt appetite upon mPFC inactivation (Supplementary Fig. 2a). A two-way repeated measures ANOVA revealed increased consumption of the sodium solution in sodium-depleted animals (main effect of state), which was driven by an increase in the frequency of licking bouts as well as by the size of these bouts (Fig. 3b, left panels). Inactivation of the mPFC, however, did not impact consumption of the sodium solution (Fig. 3b, left panels), nor of demineralized water (Fig. 3b, right panels), as the ANOVA revealed no main effect of B/M or B/M × state interaction effect.

Inactivation of the NAc (Supplementary Fig. 2b) significantly decreased sodium intake, as the two-way repeated measures ANOVA revealed a main effect of B/M on the licks of salt, which was driven by a decrease in the number of licking bouts but not by the size of these bouts (Fig. 3c, left panels). However, in a sodium-depleted state, animals still drank a substantial amount of salt, even after B/M infusion (on average 323 ± 148 s.e.m. licks in the 1h recording session). Indeed, there was a significant main effect of sodium depletion (state) on the number of sodium licking bouts, and a trend towards an effect of sodium depletion on the number of licks and bout size. Importantly, no B/M × state interaction effects were observed, indicating that the effects of sodium deprivation on salt intake were still present after NAc inactivation, although numerically more modest.

Licking for water in sodium-deprived and control rats also decreased upon infusion of B/M into the NAc, as a significant main effect of B/M was observed (Fig. 3c, right panels). In contrast to licking for salt, water consumption was almost fully abolished in both groups of rats (on average 4 ± 1-2 s.e.m. licks in the 1h recording session), without a main effect of state or B/M × state interaction effect. Collectively, these data show that inactivation of the NAc decreases intake of salt, but not as strongly as for water.

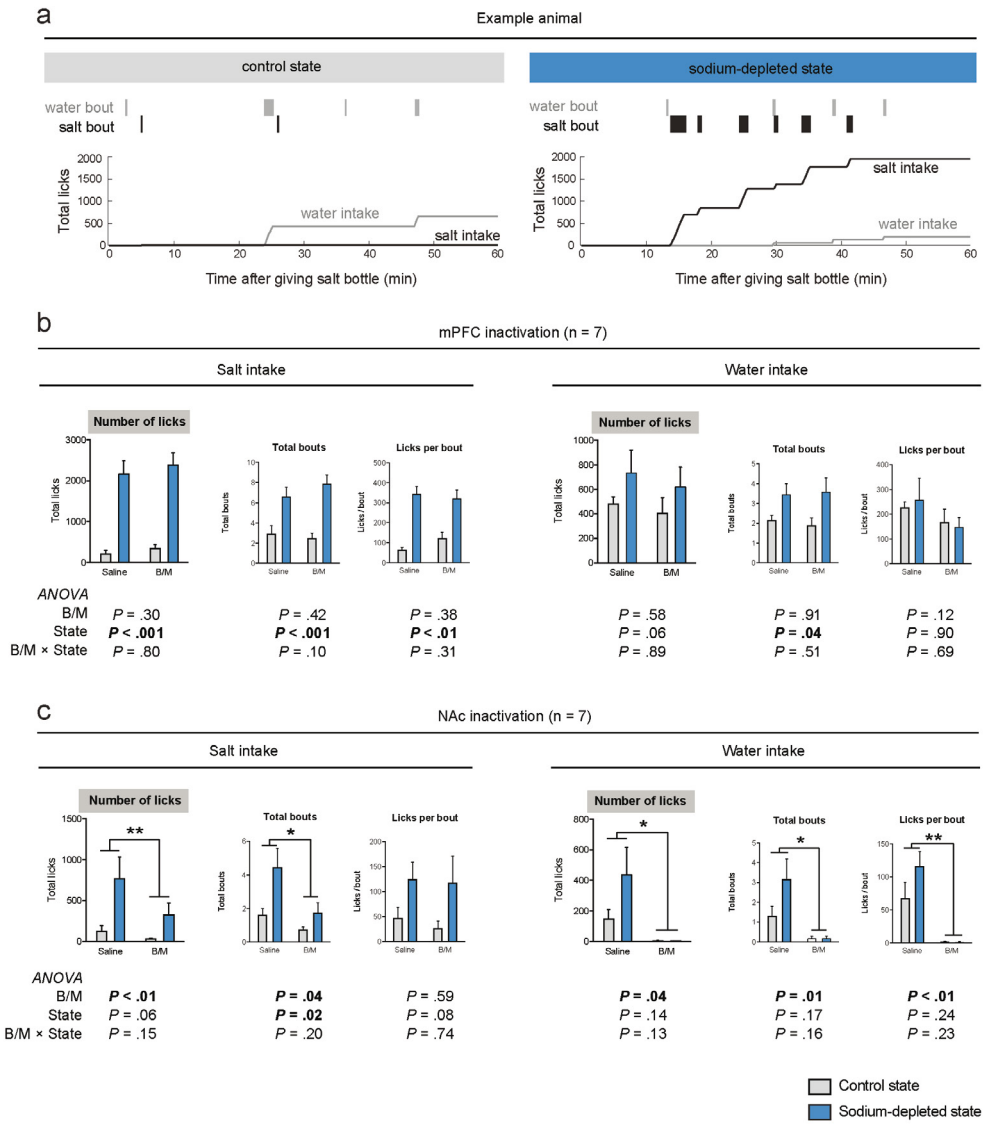


Figure 3 Effects of pharmacological inactivation of VTA target regions on salt appetite.
a) Microstructure analysis of licking behavior in an example animal once in a control state (left) and once in a sodium-depleted state (right). At time = 0 min the salt bottle was given back to the animal and its drinking behavior was analyzed as number of licks (grey line for water intake, black line for salt intake). On the upper part of the graph, bout analyses for salt and water intake shows frequency and size of the bouts. **b)** Effect of mPFC inactivation on salt intake (left) and water intake (right). No main effect of mPFC inactivation by baclofen and muscimol (B/M) of interaction effect was detected. **c)** Effect of NAc inactivation on salt intake (left) and water intake (right). Inactivation of the NAc decreased sodium intake, which was driven by a decrease in the number of licking bouts. A significant main effect of state was detected for the number of sodium licking bouts, and a trend towards an effect of sodium depletion on the number of licks and bout size. No B/M × state interaction effects were observed. Inactivation of the NAc also abolished water consumption, as a main effect of state was found on the number of water licks, driven by effects on the number of bouts and licks per bout. ** P < 0.01, * P < 0.05

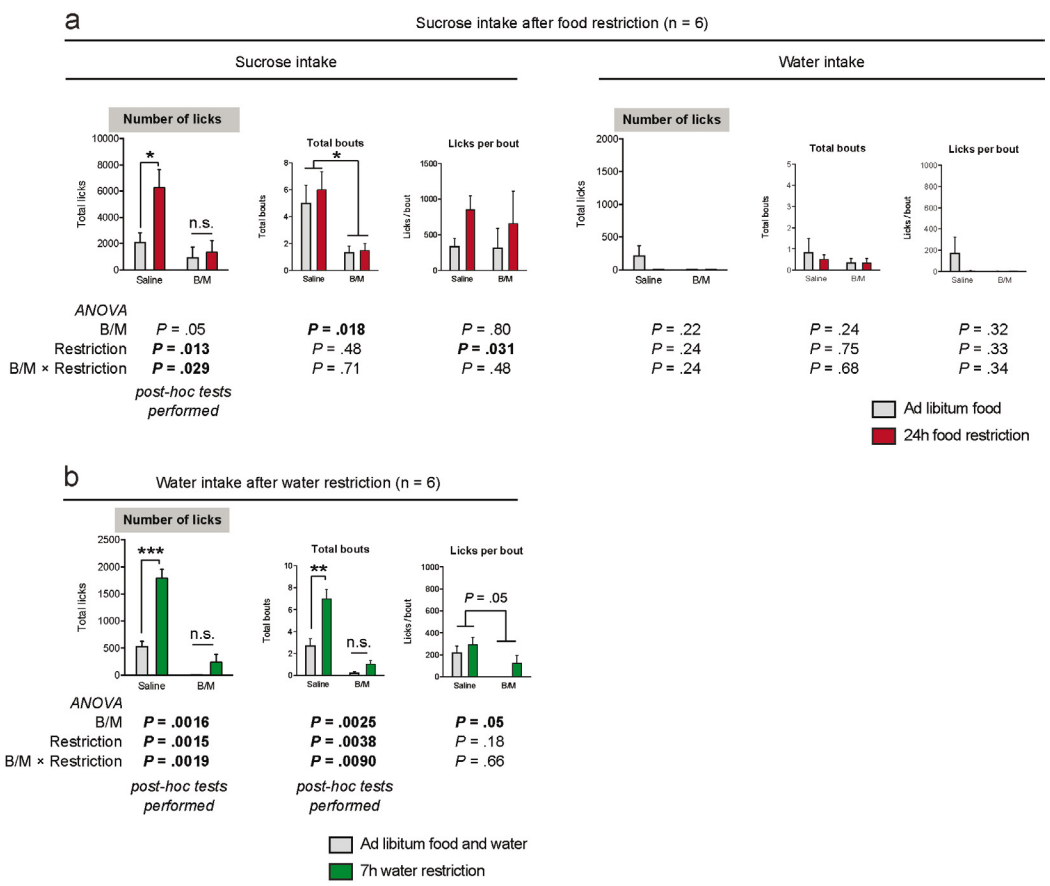


Figure 4 NAc inactivation reduced sucrose and water intake.
a) Sucrose (left) and water (right) intake was analyzed when animals were in a food restricted state (red) or in an ad libitum fed state (grey). A significant B/M × food restriction interaction effect on the number of licks for sucrose was found. Post-hoc tests Sidak's test revealed a significant increase in the number of licks due to food restriction after saline infusion (t_5 = 4.77, P = 0.010), but not after B/M infusion (t_5 = 0.48, P = 0.877). A decrease in the number of licking bouts was found, which indicates that the interaction effect was mainly driven by a decrease in motivation for sucrose. Food restriction increased the number of licks for sucrose, driven by an increase in licks per bout. Water intake was extremely low and no significant effects could be detected on water licking behavior.
b) Water licking behavior was analyzed after water restriction. Water restriction increased the number of water licks, driven by an increase in number of licking bouts. NAc inactivation decreased overall water intake, as a significant main effect of B/M on the number of licks and the number of bouts were detected, as well as a trend towards a main effect of licks per bout. A significant B/M × water restriction interaction effect was observed of the number of licks for water which was driven by an increase in licking after saline infusion (t_5 = 10.29, P = 0.0003) but not after B/M infusion (t_5 = 1.87, P = 0.23) as revealed by post-hoc Sidak's tests. There was also a significant B/M × water restriction interaction effect on the number of bouts, driven by an increase in bouts after saline (t_5 = 7.24, P = 0.0016), but not B/M (t_5 = 1.39, P = 0.40) infusion. *** P < 0.001, ** P < 0.01, * P < 0.05

NAc inactivation abolished sucrose and water intake, even during hunger and thirst
Since we observed that NAc inactivation almost fully abolished water, but not salt intake, we next examined the effects of NAc inactivation on food intake during hunger, and later also assessed the effects of NAc inactivation on water intake during thirst. We used the same experimental design as we had used to assess salt appetite, but instead monitored the intake of a 5% sucrose solution in the home cage after food restriction. As such, animals had the choice between a bottle of sucrose (which was delivered to the animal right after the infusion) and a bottle of tap water (which was already present in the home cage of the animals). Animals in the control state, who were *ad libitum* fed, had access to regular chow before and during the experiment.

We observed a significant B/M × food restriction interaction effect on the number of licks the animals made for sucrose (Fig. 4a, left panels). Post-hoc tests indicated that this was driven by a significant increase in the number of licks for sucrose upon food restriction after saline infusion, but not after B/M infusion. This effect seemed mainly driven by a decrease in the number of licking bouts, as we observed a main effect of B/M on this parameter, but not on the number of licks per bout. In contrast to baseline sucrose consumption, the total intake of water was extremely low (Fig. 4a, right panels), perhaps because the animals' water homeostasis was relatively normal (compared to after sodium deprivation) and the animals had continuous access to water. Together, these data demonstrate that NAc inactivation reduced consumption of sucrose, and that this is independent of the energy balance of the animal.

Similar effects of NAc inactivation were observed on the intake of water during thirst (Fig. 4b). Water restriction increased the consumption of water (main effect of restriction on licks and the number of licking bouts), and B/M infusion into the NAc decreased overall water intake (significant main effect of B/M on number of licks and number of bouts; trend towards a main effect on the licks per bout). Furthermore, a significant B/M × water restriction interaction effect was observed on the number of licks for water and the number of licking bouts, which was driven by an increase in licking after saline infusion but not after B/M infusion. This indicates that pharmacological inactivation of the NAc abolished water intake, even when animals were thirsty.

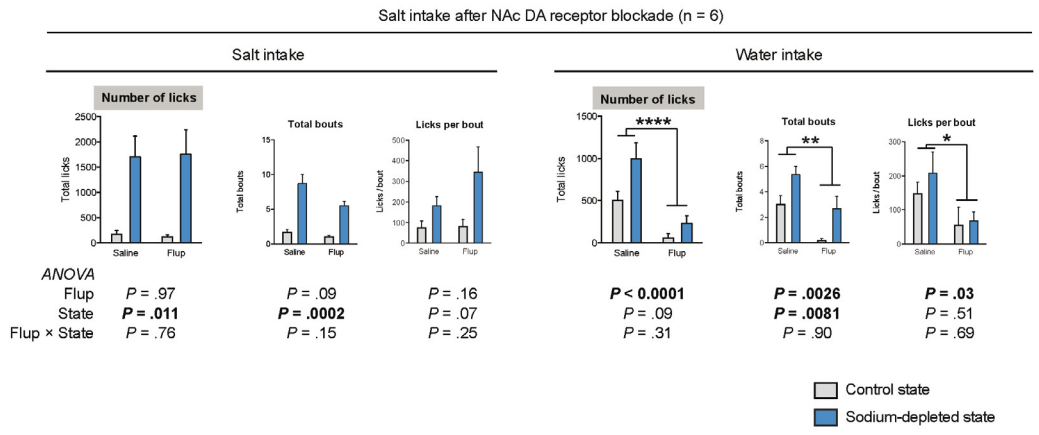


Figure 5 Effects of DA receptor blockade in the NAc on salt appetite. The effects of infusion of the DA receptor antagonist α-flupenthixol (Flup) on salt (left) and demineralized water (right) intake in rats in a sodium-depleted (blue) and control (grey) state. Infusion of α-flupenthixol did not affect salt intake, nor the number of bouts or the number of licks per bout. Water intake was significantly decreased by infusion of the DA receptor antagonist, driven by decreases in both the number of bouts and licks per bout.
**** P < 0.0001, ** P < 0.01, * P < 0.05

DA receptor antagonism in the nucleus accumbens does not alter salt appetite

We next repeated the salt intake experiment in these animals, but now infused a high dose (25 µg/side) of the DA receptor antagonist α-flupenthixol into the NAc, to study the importance of DAergic neurotransmission in the NAc for salt appetite. We observed that α-flupenthixol infusion did not affect salt intake, nor did it affect the number of licking bouts or licks per bout (Fig. 5, left panels). However, we did observe a significant effect of α-flupenthixol infusion on water intake (driven by both a decrease in the number of licking bouts and the size of these licking bouts; Fig. 5, right panels). These data suggest that the suppressing effects of pharmacological inactivation of the NAc on salt intake under sodium-depleted conditions is not driven by DAergic neurotransmission.

Discussion

In our study, we demonstrated that VTA DA neurons in rats encoded the appreciation of a salty solution, dependent on the homeostatic state of the animal. As such, tasting salt under normal circumstances resulted in sub-baseline levels of DA neuron activity, in accordance with this solution being considered aversive. Conversely, salt tasting after sodium depletion evoked vigorous licking for the solution, along with peaks in DA neuron activity that were even larger than the peaks previously observed during sucrose tasting. This finding is consistent with a recent study that demonstrated altered DA release in the nucleus accumbens shell in response to a NaCl solution after salt deprivation[16]. Another study[17] recently showed that sodium deprivation did not affect baseline activity of VTA DA neurons, *ex vivo* nor *in vivo,* in accordance with our finding that *c-Fos* expression was not altered in midbrain DA nuclei after sodium deprivation. These findings suggest that sodium deprivation does not simply disinhibit the whole DA system, but that the changes in DA neuron activity are dependent on presentation of the salient salt solution.

Several studies have suggested that during free-intake paradigms, the frequency of licking bouts (i.e., how often the animal initiates drinking) is a measure of incentive salience, or the motivation to obtain reward, whereas the bout size (i.e., the length of a drinking period) informs about the hedonic impact, or appreciation of reward[18-21]. We therefore performed a microstructure analysis of salt licking behavior after pharmacological inactivation of the NAc and demonstrated that the decrease in salt intake was driven by a reduction in the number of licking bouts, but not the size of these bouts. This finding was replicated in the sucrose intake experiment, where we observed that NAc inactivation reduced the number of sucrose licking bouts, but not the number of licks within these bouts. Together, this suggests that the motivational aspect of salt appetite is reduced by inactivation of the NAc, just as is the motivation for calories. In contrast to these findings, NAc inactivation did have a significant effect on the size of the licking bouts for water, a liquid that has a neutral taste, suggesting that the absence of an effect of NAc activation on the sucrose and salt licking bout size is related to its taste, and that taste acts as the conditioned stimulus for the homeostatic need. In fact, NAc inactivation almost fully suppressed water intake in all of the experiments, even when the animals were thirsty. The observation that this was not the case for salt and sucrose consumption suggests that the attenuated water intake was not the result of a general behavioural impairment, for example because of motor deficits. In contrast to inactivation of the NAc, we found no effects of inactivation of the mPFC on salt appetite.

After pharmacological blockade of NAc DA receptors using α-flupenthixol, we observed no effects on salt appetite. Interestingly, we did again observe an effect of the DA receptor antagonist on water intake during this experiment, just as after inactivation of the NAc with B/M. This suggests that the motivation for salt does not require accumbens DA, but that this is not necessarily the case for other types of motivation. The lack of effect of NAc DA receptor blockade on salt appetite is somewhat surprising, given that mesoaccumbens DA is considered a driving force behind motivation for rewards[22,23]. Furthermore, previous studies have reported alterations in the mesolimbic dopamine system and its inputs after sodium deprivation, both morphologically[15] and functionally[16]. Our findings are not necessarily conflicting with these data, as we also showed that VTA DA neuron dynamics during salt and sucrose tasting are dependent on the sodium balance of the animal. This suggests altered reward processing after sodium deprivation, which may logically also affect downstream DA release and hence morphological and structural changes to downstream areas. However, we do show that mesolimbic DA neurotransmission is not *necessary* for the behavioral expression of salt appetite.

That said, the finding that a switch in salt appreciation upon a change in body sodium levels is encoded by VTA DA neurons, but that blockade of DA receptors in one of its most important downstream regions does not hamper salt appetite may seem counterintuitive. A possible explanation is that behavioral adaptation to a shortage in sodium is so crucial, since it can be a prerequisite for survival, that it is redundantly coded in the brain, and thus relies on a variety of brain regions. For example, the VTA also projects to the subthalamic nucleus, which projects to the substantia nigra and the ventral pallidum, which again sends efferents to the substantia nigra, lateral hypothalamus, lateral pre-optic area, pedunculopontine nucleus, and brainstem[24]. All these regions form a complex network of the ventral basal ganglia, which could function as backup for dysfunction of the NAc. Furthermore, the VTA is known to directly project to the lateral hypothalamus, also a key region of the reward system, which forms a neural circuit with the parabrachial nucleus and the nucleus of the solitary tract, which has shown to be involved in the sensory and motor aspects of feeding[25,26].

Altogether, we have used a multidisciplinary approach, including c-Fos immunohistochemistry, fiber photometry and behavioral pharmacology, to assess the role of the mesocorticolimbic DA system in salt appetite. We have substantiated findings from earlier studies regarding the role of VTA neurons in salt appetite, and provide novel insights into the role of its target regions in this behavior. We show that the NAc, but not mPFC, is essential for the behavioral expression of salt appetite by mediating its motivational, but not hedonic, component. This role of the NAc in salt appetite is independent of DA, although we show that DA neurons themselves do encode the appreciation of salt through reward prediction error.

Methods

Animals

All experiments were approved by the Animal Ethics Committee of the Utrecht University, and were conducted in agreement with Dutch (Wet op de Dierproeven, revised 2014) and European regulations (Guideline 86/609/EEC; Directive 2010/63/EU).

A total of 46 male rats were used in the experiments. Male Long-Evans rats (Rj:Orl; Janvier Labs, France) were used for the micro-infusion experiments (*n* = 14), male Wistar rats (Crl:WU; Charles River, Germany) were used for c-Fos analysis (*n* = 22), and TH::Cre transgenic rats (bred in-house by crossing heterozygous TH::Cre+/- male rats with wild type Rj:Orl mates) were used for fiber photometry (*n* = 6 TH::Cre+ injected with DIO-GCaMP6s and, *n* = 4 TH::Cre- injected with eYFP). All rats weighted ~250 g at the start of the experiments, were individually housed under controlled temperature (20°C) conditions, with a 12h light/dark cycle (lights off at 7:00 a.m.), and received a wood block as cage enrichment. When not being tested, animals had *ad libitum* access to demineralized water and a 0.45 M sodium chloride solution (or a 5% sucrose solution, prior to the sucrose intake experiment) and standard chow (Special Diet Service, UK) in the home cage. Preceding sodium intake test days, animals were salt deprived, during which they only had access to demineralized water and a sodium-deficient chow (Teklad Custom Diet, Envigo). Preceding the sucrose intake test days, animals were food restricted, during which they had no access to chow or the sucrose solution for 24 h.

Surgeries
Anesthesia was induced using a mixture of 0.315 mg/kg fentanyl and 10 mg/kg fluanisone (Hypnorm, Janssen Pharmaceutica, Belgium) that was injected intramuscular. Animals were placed in a stereotaxic apparatus (David Kopf Instruments, USA) and an incision was made along the skull midline.

For fiber photometry experiments, the same surgical procedure was applied as described previously[10]. In brief, TH::Cre rats were injected with 1 µl of AAV5-FLEX-hSyn-GCaMP6s (University of Pennsylvania Vector Core) at a titer of 1×10^{12} particles/ml unilaterally into the right VTA (−5.40 mm AP, ±2.20 mm ML from Bregma, at an angle of 10°, and −8.90 mm DV from the skull). A 400 µm implantable fiber was lowered to 0.1 mm above the injection site and attached with dental cement.

For micro-infusion experiments, 26-gauge stainless steel guide cannulas (Plastics One, USA) were implanted above the NAc (two single cannulas; +1.20 mm anteroposterior (AP), ±2.80 mm mediolateral (ML) from Bregma, at an angle of 10°, and the guide was lowered to -6.80 mm dorsoventral (DV) from the skull) or the mPFC (one double cannula with a width of 1.2 mm; +3.20 mm AP, ±0.60 mm ML from Bregma, and the guide was lowered to -2.60 mm DV from the skull). Cannulas were secured to the skull with screws and dental cement, and dummy injectors were placed inside the cannulas to prevent blockage. Single injectors for the NAc protruded 0.5 mm beyond the guides (targeting -7.30 mm DV from the skull) and double injectors for the mPFC protruded 1 mm beyond the guides (targeting -3.50 mm DV from the skull).

To prevent dehydration of the rats, they were given 10 mL of saline subcutaneous (s.c.) after surgery. Starting on the day of surgery, rats were given carprofen as analgesia (s.c. injection of 5 mg/kg per day for 3 days). All rats were allowed to recover from surgery for at least 7 days before behavioral testing began.

Sodium deprivation
Before the sodium deprivation procedure, all cages were cleaned to prevent the rats from repleting their sodium levels by eating their own feces. Sodium depletion was induced by an s.c. injection of the diuretic drug furosemide (20 mg/kg dissolved in sterile H_2O, given in 2 injections of 10 mg/kg 1 hour apart). Control animals received s.c. saline injections. In the 24h that followed, sodium-depleted animals received sodium-free chow, and control animals received regular chow. In the first three hours after the first furosemide injection, animals had no access to water, to confirm success of the procedure by observing a body weight loss. After these 3 hours, all animals received demineralized water, which was especially heavily consumed by the animals that were previously injected with furosemide. 24 hours after the first furosemide injection, animals were given a bottle containing a 0.45M NaCl solution, and intake of this solution (as well as intake of the demineralized water, which was already present in the cage) was monitored for 1h using mechanical lickometers that were present in the home cage. Animals were always tested in a counterbalanced fashion, so that half of the animals were first tested in a control state, i.e., 24 h after s.c. saline injection, and the other half in a sodium-depleted state, i.e., 24 h after s.c. furosemide injection. Drinking behavior was assessed as cumulative intake (number of licks), number of licking bouts, and licks per licking bout for both the intake of demineralized water and the 0.45M NaCl solution. A minimum of 5 licks was considered a bout, which ended when the animals did not lick for at least 1min.

***In vivo* fiber photometry**
Technical details about our fiber photometry setup have been published elsewhere[10]. In brief, animals were injected with a Cre-dependent GCaMP6s in the right VTA, and a 400 µm fiber was secured 0.1mm dorsal of the injection site. Animals were connected to a 400 µm core fiber optic patch cable through which lock-in amplified blue LED light was delivered.

Emission light was captured with a photoreceiver, digitized, and dF/F_0 values were computed with F_0 being defined as the mean of the middle 50% of values in the 30 seconds before each time point F.

Each rat was tested on the behavioral task twice, once in a salt-depleted state and once in a control state, and the task was conducted in operant conditioning chambers (MedPC Inc., USA). The chambers were equipped with one optical lickometer (delivering both solutions through the same spout), and on the other side of the chamber a house light and auditory tone generator. All animals were food restricted for 24h before the measurement, to increase the motivation for sucrose (making sure the animals lick during every trial).

In the task, a 5-second tone initiated the trial, and the first lick after tone offset triggered the fluid pump, which delivered a droplet of the solution over a period of 5s. If the animal did not make a lick within 5s after tone offset, no reward was obtained and the inter-trial interval of 30s commenced. If the animal did make a lick within 5s after tone offset, the pump delivered a 0.88M sucrose solution in 75% of the trials and a 0.30M NaCl solution in 25% of the trials (in random order). After the 5-second liquid delivery, a 30-second inter-trial interval separated the current trial from the onset of the next trial**.** No cue lights were used and the house light was turned on continuously to prevent the signal to be contaminated by lights from the environment. Individual trial responses were time-locked to the 5-second tone that started the trial and mean dF/F of trial responses to sucrose and of trial responses to sodium was calculated. The number of licks during the trials was assessed using the lickometers that were monitored by medPC software. The task continued until the animal had made at least 80 trials.

Microinfusions
For the infusion experiments, n = 7 (NAc) and n = 8 (mPFC) rats were used. Animals were habituated to the infusion procedure by infusing saline (0.5 µl/side) the day before the first experiment. Rats were brought in a salt depleted state or in a control state, as described in the paragraph above, and 24 h later they received infusions with saline (1 µl/side for the NAcc, 0.5 µl/side for the mPFC) or a mixture of baclofen (1nmol; Sigma-Aldrich, The Netherlands) and muscimol (0.1 nmol; Sigma-Aldrich, The Netherlands) dissolved in saline (1 µl/side for the NAcc, 0.5 µl/side for the mPFC). Furosemide vs saline injections and baclofen-muscimol vs saline infusions were performed in a Latin Square repeated measures design. Drugs were infused at a rate of 0.5 µl/min, and the injectors were left in place for an additional 30 s after the infusion was complete to allow for diffusion of saline/baclofen-muscimol into the brain. After the infusion procedure, animals were placed back in their home cage, and a bottle with a 0.45 NaCl solution was given 5 minutes later. In the dopamine receptor antagonist infusion experiment, we used the same experimental procedure as in the pharmacological inactivation experiments, except that 25 µg of cis-(Z)-α-flupenthixol dihydrochloride (Sigma-Aldrich, The Netherlands) was infused, dissolved in 0.5 µl saline.

c-Fos analysis
For c-Fos analysis, 11 animals were brought in a salt-deprived state as described in the procedure above, and 11 animals were used as control animals. 48 hours after the first furosemide injection, all 22 rats received an i.p. injection of sodium pentobarbital and perfused with phosphate-buffered saline (PBS) followed by 4% paraformaldehyde (PFA) in PBS. After extraction, brains were post-fixated in 4% PFA in PBS at 4 °C for 24 h, and stored in a 30% sucrose in PBS solution at 4 °C. For immunohistochemical quantification of the number of activated neurons, brains were stained for the immediate early gene c-Fos. Brain slices (50 µm) were blocked in 3% normal goat serum (NGS) and 0.5% Triton-X-100 in PBS. Slices were incubated overnight in primary antibody rabbit anti-c-Fos (1:1000, Cell Signaling) in 3% NGS in PBS at room temperature. Subsequently, slices were incubated for 2 h in biotinilated antibody goat anti-rabbit (1:200, Vector labs) in 3% NGS in PBS, and

afterwards in Biotin/Avidin (1:1000, Vectastain) in PBS for 1 h. This complex was visualized by exposing the slices for 5 min to a solution of liquid DAB (3,3'-Diaminobenzidine, Dako) and 10% nickel ammonium sulphate. All sections were dehydrated using increasing series of ethanol, cleared in xylene and coverslipped with Entallan (Merck Millipore). Sections were photographed by a brightfield microscope with a 10X lense (AxioImager M2). Slices comprising the VTA were manually aligned in illustrator, and ImageJ (Version 1.48 v) was used to extract the coordinates of c-Fos positive neurons by applying a bandpass filter over the Fourier-transformed image, followed by a search for maximum intensity points. Heatmaps of c-Fos expression were generated based on the coordinates of the c-Fos positive cells using MATLAB (The MathWorks Inc., version R2014a).

Exclusion criteria
One animal was excluded from c-Fos analysis because brain slices were not of sufficient quality. One additional animal was excluded from c-Fos analysis because the brain slices did not show any expression. One animal was excluded from the sucrose intake (after food restriction) experiment, water intake (after water deprivation) experiment and dopamine receptor antagonist infusion experiment, because it was suspected to develop diabetes (it drank excessive amounts of water and sucrose water, and the bedding was continuously wet).

Data availability
The datasets generated during the current study are available from the corresponding author on reasonable request.

Data analysis and statistics
Data analysis was performed with MATLAB, statistical analysis using GraphPad Prism (GraphPad Software Inc., version 6.0). Statistical comparisons were made using a two-tailed t-test for a single comparison and a (repeated measures) ANOVA was used for multiple comparisons, followed by a t-test with Šidák's multiple comparisons correction when a significant interaction effect ($P < .05$) was found between the two factors of the ANOVA. Bar graphs represent the mean ± standard error of the mean. In all figures: ns not significant, $^{\#}P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

References

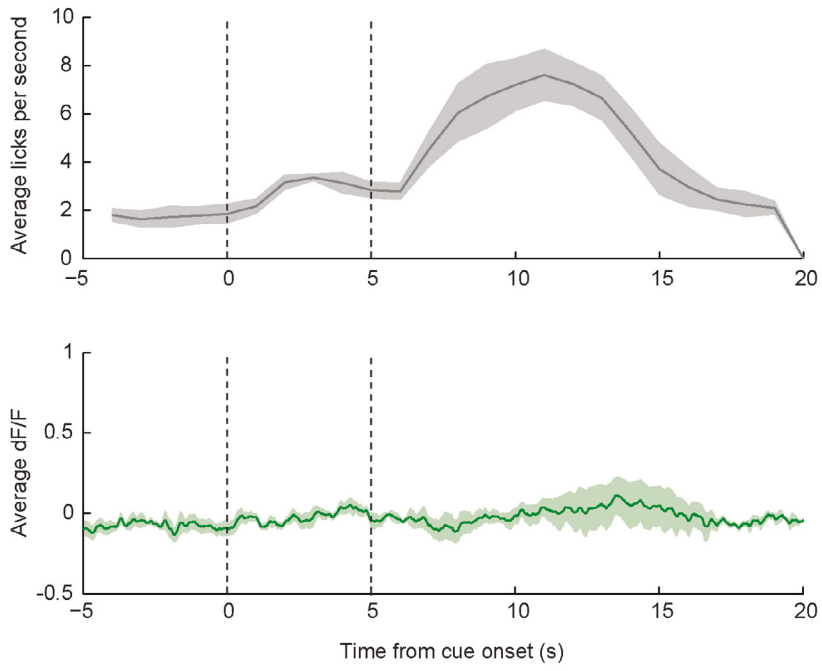
1. Richter, C. Salt appetite of mammals: its dependence on instinct and metabolism. L'Instinct dans le comportement des animaux et de l'homme 577 (1956).
2. Roitman, M. F., Schafe, G. E., Thiele, T. E. & Bernstein, I. L. Dopamine and Sodium Appetite: Antagonists Suppress Sham Drinking of NaCl Solutions in the Rat. Beh. Neurosci. 111, 606-611 (1997).
3. Berridge, K. C., Flynn, F. W., Schulkin, J. & Grill, H. J. Sodium Depletion Enhances Salt Palatability in Rats. Beh. Neurosci. 98, 632-660 (1984).
4. Berridge, K. C. Palatability Shift of a Salt-Associated Incentive During Sodium Depletion. The Quarterly J. of Exp. Psychol. 41B, 121-138 (1989).
5. Menani, J. V., De Luca Jr, L. A. & Johnson, A. K. Lateral parabrachial nucleus serotonergic mechanisms and salt appetite induced by sodium depletion. Am. J. Physiol.-Reg I. 274, R555-R560 (1998).
6. Geerling, J. C. & Loewy, A. D. Aldosterone sensitive neurons in the nucleus of the solitary tract: Efferent projections. J. of Compar. Neurol. 497, 223-250 (2006).
7. Jarvie, B. C. & Palmiter, R. D. HSD2 neurons in the hindbrain drive sodium appetite. Nat. Neurosci. (2016).
8. Geerling, J. C. & Loewy, A. D. Central regulation of sodium appetite. Exp. Physiol. 93, 177-209 (2008).
9. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. Science 275 (1997).
10. Verharen, J. P. H. et al. A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. Nat. comm. 9 (2018).
11. Wolf, G. Hypothalamic regulation of sodium intake: relations to preoptic and tegmental function. Am. J. of Psychol. 213, 1433-1438 (1967).
12. Stricker, E. M. & Zigmond, M. J. Effects on homeostatis of intraventricular injections of 6-hydroxydopamine in rats. J. of Comp. and Physiol. Psych. 86, 973-994 (1974).
13. Lucas, L. R., Grillo, C. A. & McEwen, B. S. Salt appetite in sodium-depleted or sodium-replete conditions: possible role of opioid receptors. Neuroendocrinology 85, 139-147 (2007).
14. Roitman, M. F., Patterson, T. A., Sakai, R. R., Bernstein, I. L. & Figlewicz, D. P. Sodium depletion and aldosterone decrease dopamine transporter activity in nucleus accumbens but not striatum. Am. J. Physiol. 276, 1339-1345 (1999).
15. Roitman, M. F., Na, E., Anderson, G., Jones, T. A. & Bernstein, I. L. Induction of a Salt Appetite Alters Dendritic Morphology in Nucleus Accumbens and Sensitizes Rats to Amphetamine. J. of neurosci. 22, 1-5 (2002).
16. Fortin, S. M. & Roitman, M. F. Challenges to body fluid homeostasis differentially recruit phasic dopamine signaling in a taste-selective manner. J. of Neurosci. (2018).
17. Sandhu, E. C. et al. Phasic Stimulation of Midbrain Dopamine Neuron Activity Reduces Salt Consumption. eNeuro 5 (2018).
18. Davis, J. D. The Microstructure of Ingestive Behavior. Ann. of the N. Y. Ac. of Sci. 575, 106-119 (1989).
19. Higgs, S. & Cooper, S. J. Evidence for early opioid modulation of licking responses to sucrose and Intralipid: a microstructural analysis in the rat. Psychopharmacology 139, 342-355 (1998).
20. D'Aquila, P. S. Dopamine on D2-like receptors "reboosts" dopamine D1-like receptor-mediated behavioural activation in rats licking for sucrose. Neuropharmacology 58, 1085-1096 (2010).
21. Ostlund, S. B., Kosheleff, A., Maidment, N. T. & Murphy, N. P. Decreased consumption of sweet fluids in mu opioid receptor knockout mice: a microstructural analysis of licking behavior. Psychopharmacology (Berl) 229 (2013).
22. Cools, R. Role of dopamine in the motivational and cognitive control of behavior. Neuroscientist 14, 381-395 (2008).
23. Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. Neuron 76, 470-485 (2012).
24. Humphries, M. D. & Prescott, T. J. The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. Prog. Neurobiol. 90, 385-417 (2010).
25. Grill, H. J. Leptin and the systems neuroscience of meal size control. Frontiers in neuroendocrinol. 31, 61-78 (2010).
26. Stice, E., Figlewicz, D. P., Gosnell, B. A., Levine, A. S. & Pratt, W. E. The contribution of brain reward circuits to the obesity epidemic. Neurosci. & Biobeh. Rev. 37, 2047-2058 (2013).

ACKNOWLEDGEMENTS
This work was supported by the European Union Seventh Framework Programme under grant agreement number 607310 (*Nudge-IT*), and the Netherlands Organisation for Scientific Research (NWO) under project numbers 912.14.093 (*Shining light on loss of control*).

AUTHOR CONTRIBUTIONS
J.P.H.V., T.J.M.R, L.J.M.J.V. and R.A.H.A designed the experiments. J.P.H.V., T.J.M.R, S.M. and M.C.M.L. performed the experiments. J.P.H.V. and T.J.M.R analyzed the data. J.P.H.V., T.J.M.R, L.J.M.J.V. and R.A.H.A wrote the paper with input from the other authors.

COMPETING INTEREST
The authors declare no competing interests.

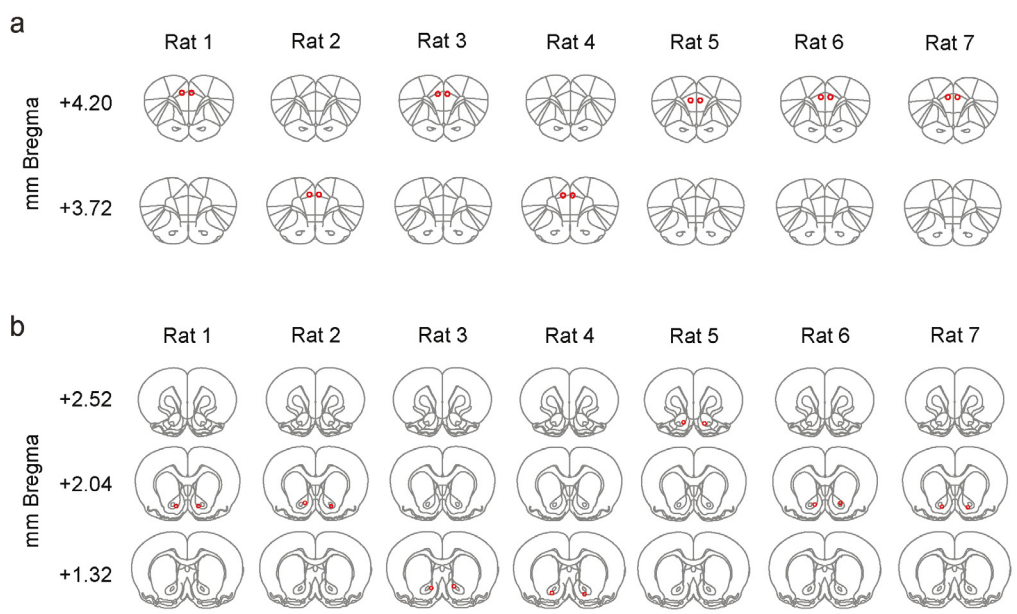
SUPPLEMENTARY FIGURE 1



In vivo fiber photometry of VTA neurons from animals injected with a YFP control fluorophore.

An activity-independent control fluorophore was injected into the VTA of control animals (n=4) and in vivo fiber photometry indicated no changes in fluorescent activity in these controls (lower panel). Upper panel shows the average licking rate of the animals. Line and shading represent mean and standard error of the mean, respectively.

SUPPLEMENTARY FIGURE 2



Histological verification of guide cannula placement.

Correct placement of the guide cannulas used for local infusions was verified for all animals in which the mPFC (a) or the NAc (b) was targeted. Rat 1 from (b) was excluded from figures 4 and 5 because it developed diabetes.

CHAPTER 9

Insensitivity to monetary losses in anorexia nervosa patients

Jeroen P.H. Verharen
Unna N. Danner
Sabrina Schröder
Emmeke Aarts
Annemarie A. van Elburg
Roger A.H. Adan

Manuscript in preparation

Highlights

- We analyzed data from a large cohort of anorexia nervosa patients performing the Iowa Gambling task
- Fitting the data to a prospect utility function model demonstrates that anorexia nervosa patients have a reduced sensitivity to monetary losses
- This suggests that anorexia nervosa patients have impairments in value-based decision making

Techniques

Computational modeling

CHAPTER 9

Anorexia nervosa patients consistently demonstrate impairments in laboratory measures of value-based learning and decision making. The mechanisms that underlie these changes have been elusive, but recent data suggest alterations in the dopamine system related to reward-based decision making. Here, we fit data of anorexia nervosa patients and healthy controls performing the Iowa Gambling task to a computational model based on prospect utility theory, and show that anorexia nervosa patients, in contrast to healthy controls, do not exhibit loss aversive behavior. This finding provides fundamental insights into the decision making capacity of anorexia nervosa patients, suggesting alterations in the mechanisms involved in value processing.

Introduction

A growing body of evidence suggests that anorexia nervosa (AN) patients have impairments in value-based learning and decision making[1-8]. This is not only inferred from the clinical presentation of the disease, which includes inflexibility and distorted goal pursuit[6,9], but also from performance in several standardized laboratory tests for decision making. For example, AN patients show impairments in set shifting[10,11], show increased capacity to delay reward[12], and demonstrate reduced problem solving capacity[13] (for a systematic review see ref. 14). In line with these findings, alterations in the dopamine system, an important hub for reward-based learning and decision making[15,16], have been reported in AN patients[17,18]. For example, reward prediction errors, neural signals encoded by midbrain dopamine neurons during the delivery of unexpected reward or punishment, have been shown to be elevated in AN patients[19].

One way to assess decision making is through the Iowa gambling task (IGT)[20-22]. The IGT measures behavioral responses to monetary gains and losses by letting participants choose between four decks of cards that each differ in the amount of money one can win or lose per card (Fig. 1a,b). Two decks give a high financial yield ($100/card) and therefore seem advantageous. However, these decks also lead to an occasional loss of a large sum of money, which results in a net loss when chosen exclusively. The two other decks give a more modest payout ($50/card) but also yield a lower amount of losses, and are therefore the advantageous option on the long term. In order to choose the profitable decks and thereby win the highest amount of money at the end of the session, one must explore each of the choice options, integrate the profits and losses associated with each of the decks into an expected reward value, and make decisions based hereon. By assessing choice behavior of participants and comparing this between different groups, one may infer alterations in decision making behavior under pathophysiological conditions, including AN. Indeed, over the years, many studies have attempted to demonstrate decision making deficits in AN patients by utilizing the IGT. A recent systematic meta-analysis that compared those studies showed a consistent lower IGT net score in symptomatic AN patients as compared to healthy controls[23], providing further evidence for impairments in value-based decision making in AN.

IGT performance is usually assessed by plotting the fraction of choices for the advantageous decks over the session. Although this metric is useful to assess whether learning takes place within a session, this measure does not directly inform about which of the underlying component processes is altered. In recent years, several attempts have been made to extract the different components of value-based decision making from the IGT data by means of computational trial-by-trial analyses[24]. One study systematically compared a wide range of reinforcement learning models in their ability to explain choice behavior in the IGT and demonstrated that a model based on prospect utility theory was superior in this aspect[25]. This theory[26-28] states that people are no perfect rational decision making agents, in the sense that under uncertainty, the subjective experience of reward is not linearly proportional to the actual received reward (Fig. 1c, inset). Rather, subjective reward is thought to be concave to the actual reward (and convex for losses), so that winning $200 has a lower impact on behavior than winning $100 twice. Furthermore, the prospect utility value function is asymmetric for negative and positive values, so that, for most people, losses weigh heavier than gains in terms of their impact on choice behavior. In other words, most people are loss averse.

Here, we use computational trial-by-trial analysis of IGT data of a reasonably large cohort of AN patients and healthy controls in an attempt to elucidate the basic computational processes that underlie the impaired performance of AN patients in the IGT. By fitting the data to a large set of reinforcement learning models, we confirm that prospect utility is the best descriptor of behavior in the IGT. After comparing the computational model coefficients between patients and controls, we demonstrate that AN patients, in contrast to healthy controls, do not exhibit loss aversive behavior.

Methods

Participants

For this study, n = 115 participants were included, from which n = 60 were diagnosed with AN and were symptomatic at the time of testing (Table 1). The second cohort included 216 participants, all of whom were diagnosed with AN and symptomatic at the time of testing. Disease classification was performed by eating disorder experts (all medical doctors) according to the DSM-V criteria. All participants were recruited at the Altrecht Clinic for Eating Disorders Rintveld, a specialized center for eating disorders in Zeist, The Netherlands.

Task

A computerized version of the original IGT[20] was used to assess decision-making ability. The IGT simulates real-life decision making under uncertain circumstances with a conflict between immediate reward and delayed punishment so that participants have to make advantageous choices. Participants are instructed to maximize their profit by choosing one card at a time from one of four card decks. After each choice (100 in total), a specific amount of money is awarded while at certain times, the participant also loses a fixed amount of money, resulting in a net loss. Decks A and B are considered disadvantageous, because they contain high gains but also high losses, disclosing a net value of minus 250 dollar per 10 cards. These decks have the same overall net loss but differ in frequency and degree of punishment. With smaller gains but also smaller losses, decks C and D are considered

	First cohort			Second cohort
	AN patients	Healthy controls	P-value	AN patients
Group size	60	55	-	216
% women in group	100%	100%	-	96%
Body mass index in kg/m^2	15.41 (1.90)	21.74 (2.81)	$P < 0.0001$	16.42 (2.40)
Age in years	27.28 (9.93)	24.47 (8.31)	$P = 0.1042$	22.25 (7.27)
Level of education	5.72 (0.78)	6.84 (0.37)	$P < 0.0001$	n/a

Table 1 - Demographics of participants. Numbers indicate mean (std).
P-values denote significance in an unpaired t-test.
Level of education is an arbitrary measure ranging from 1 (primary school not finished) to 7 (university)

to be advantageous in the long run, disclosing a net value of plus 250 dollar per 10 cards. These two decks also display the same overall net loss while differing in frequency and degree of punishment. Traditionally, decision making is examined by dividing the 100 trials in five blocks of 20 card choices, also referred to as the learning effect during the task. For each block, a net score is calculated by the difference in number of choices between the advantageous and disadvantageous decks: [(C+D) - (A+B)]. An impairment in decision-making ability is characterized by a lack of improvement of performance over time.

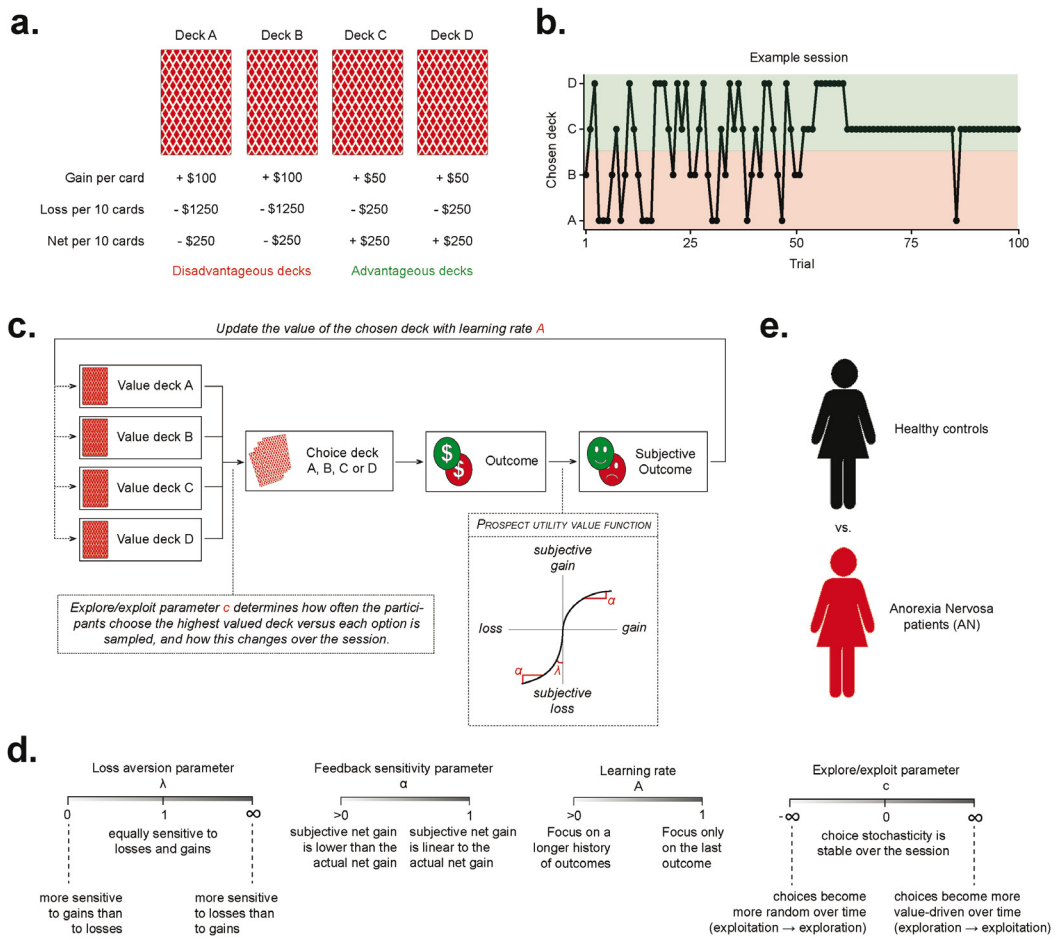


Figure 1: Iowa gambling task and computational model
a. Task design.
b. Example data of a participant that starts with an exploratory approach, but chooses more exploitative in a later stage of the session.
c. We fit a computational model to the data to mathematically dissect the different components of value-based decision making in the IGT.
d. Interpretation of model parameter values.
e. We compared healthy controls (HC) with anorexia nervosa (AN) patients.

Modeling analysis
In the modeling analysis, we tested 8 different reinforcement learning models, all as described by ref. 25 (see also refs. 29 and 30 for a comparison of IGT models). All of these models assume that participants make decisions by a process that is reiterated on every trial (Fig. 1c), and comprises 1) a utility function, that transforms the gains or losses from that trial into a net subjective value, or 'utility', 2) updating the value representations of the decks on the basis of this subjective value, and 3) make a choice between the four decks by comparing their expected values. In each of these three steps, two different types of equations were tested, so that all possible combinations of equations resulted in a total of $2^3 = 8$ models. For the utility function, an equation based on prospect utility theory was tested and an equation based on expected utility theory. For the value updating function, a delta learning rule was tested (i.e., the Rescorla-Wagner model; only updates the chosen deck based on reward prediction error) as well as a decay-reinforcement rule (which also discounts the value of a deck when it is not chosen). For the choice function, a Softmax equation was used, one which was based on the assumption that choice stochasticity (i.e., the explore/exploit trade-off parameter) was stable within a session (trial-independent choice rule), and one based on the assumption that choice stochasticity may change over a session (trial-dependent choice rule; e.g., that a participant could start with an exploratory approach, but may become more deterministic in a later stage of the task).

The trial-by-trial data of participants was fit to each of these models, and random effects model selection[31] was performed using the individual log-model evidence estimates with the function 'spm_BMS' in the Matlab toolbox SPM 12 (Wellcome Trust Centre for Neuroimaging). The model that was the best descriptor of IGT performance was model #5 (highest PXP; Table 2), which was the model based on prospect utility function, a delta learning rule and a trial-dependent choice rule. Behavior of participants in this model was described on the basis of 4 parameters: 1) loss aversion parameter *γ*, which is the steepness of the prospect utility value function for a negative outcome compared to a positive outcome,

M	Utility	Updating	Choice	Aggregate LL	Aggregate AIC	XP	PXP
1	Expected utility	Delta learning rule	Trial-dependent consistency	-12957.1	26604.2	0.000	0.000
2			Trial-independent consistency	-13082.7	26855.4	0.000	0.000
3		Decay-reinforcement learning rule	Trial-dependent consistency	-19121.5	38932.9	0.000	0.000
4			Trial-independent consistency	-12048.6	24787.3	0.114	0.114
5	**Prospect utility**	**Delta learning rule**	**Trial-dependent consistency**	**-12004.9**	**24929.8**	**0.861**	**0.861**
6			Trial-independent consistency	-13353.7	27627.4	0.000	0.000
7		Decay-reinforcement learning rule	Trial-dependent consistency	-11914.6	24749.2	0.021	0.021
8			Trial-independent consistency	-11811.1	24542,2	0.005	0.005

Table 2: Bayesian model selection (n = 115 participants) indicated that a model based on prospect utility function explained the highest amount of choices as compared to the other reinforcement learning models. Abbreviations: LL, log likelihood; AIC, Akaike Information Criterion; XP, exceedance probability; PXP, protected exceedance probability.

2) feedback sensitivity parameter *a*, which is the exponent of the prospect utility value function, 3) learning rate *A*, which reflects the strength with which a single outcome affected the value representation of the chosen deck, and 4) explore/exploit parameter *c*, indicating how choice stochasticity changed over the session. Fig. 1d shows an overview of the model parameters and the interpretation of their values.

For the 'winning' computational model (adapted from ref. 25), the value function was given by

$$u_t = \begin{cases} x_t^{\alpha} & \text{for net gains} \\ -\lambda \mid x_t \mid^{\alpha} & \text{for net losses} \end{cases}$$

Here, u is the utility on trial t, based on the net monetary outcome x_t. Here, λ and a denote loss aversion and feedback sensitivity, respectively, which are two of the free parameters in the model.

Next, the value representation of the chosen deck was updated based on the reward prediction error, which is the discrepancy between the expected outcome, $V_{\text{chosen,t-1}}$, and the actual (subjective) outcome, u_t. The reward prediction error δ_t was thus given by

$$\delta_t = u_t - V_{\text{chosen,t-1}}$$

so that the reward prediction error was positive when the net yield was higher than expected, and negative when this was lower than expected. Hence, higher-than-expected outcomes would increase the value of the chosen deck, and lower-than-expected outcomes would decrease the value of the chosen deck, so that

$$V_{\text{chosen,t}} = V_{\text{chosen,t-1}} + A \times \delta_t$$

In which A represents the learning rate, so that a low value of A indicates low learning (i.e., a focus on a longer history of outcomes, because a single outcome does not strongly affect V_{chosen}) and a high value of A indicates high learning (value is to a large extent based on the last outcome).

The values of the chosen decks V_A, V_B, V_C and V_D, were then converted into action probabilities using a Softmax rule, so that the probability of choosing, for example, deck A was given by

$$p_{A,t} = \frac{\exp(\theta \cdot V_{A,t})}{\exp(\theta \cdot V_{A,t}) + \exp(\theta \cdot V_{B,t}) + \exp(\theta \cdot V_{C,t}) + \exp(\theta \cdot V_{D,t})}$$

with the Softmax' inverse temperature θ being dependent on the trial number, so that

$$\theta = (\frac{t}{10})^c$$

To obtain reliable model parameter estimates on a population level, we used maximum a posteriori estimation. Priors were set over each of the free parameters: for λ, normpdf(2,1); for a and A, betapdf(1.3, 1.3); for c, normpdf(4,2). Subsequently, on each trial, new evidence was considered by computing the posterior probability distribution using Bayes' rule.

Statistics

All outcomes measures were tested for being normally distributed using the D'Agostino and Pearson omnibus normally test (threshold set at $P < 0.05$), after which the appropriate statistical test was performed. All computational analyses were performed with Matlab 2014a and the statistical tests were performed with GraphPad Prism 6.0. Asterisks in the figures denote statistical significance, with the following ranges: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Results

IGT performance of AN patients is altered

In total, we compared IGT performance of 60 AN patients with 55 healthy controls (Fig. 1e and Fig. 2a). In accordance with literature, we observed reduced learning over the different trial blocks in AN patients compared to control participants (Fig. 2b, left panel). After classifying AN patients into the restrictive and binge-purge subtypes, we observed visually comparable differences to the control group, although there was only a significant group × block interaction effect in the binge-purge subtype group compared to controls. However, a two-way ANOVA performed on the two AN subtypes separately revealed no significant differences between the two patient groups (group effect, $P = 0.83$, group × block interaction effect, $P = 0.15$).

AN patients exhibit decreased loss aversion

After fitting the model to the data and estimating the model parameter values, we observed a significant decrease in the estimate of loss aversion parameter λ in AN patients as compared to healthy controls (Fig. 2c). Furthermore, we performed individual one-sample statistical tests on the two groups to assess whether their λ estimates were significantly higher than 1, which would be indicative of a stronger impact of losses than wins on behavior, as would be expected based on literature[32,33]. Indeed, the estimate of λ for the controls was significantly different from 1 (Wilcoxon signed rank test, $P = 0.0002$), but this was not the case for AN patients (one-sample t-test, $P = 0.8352$). This indicates that AN patients were not loss averse, in contrast to healthy controls. No significant differences were found between AN patients and healthy controls on the estimates of feedback sensitivity parameter a, learning rate A or stochasticity parameter c.

We observed no significant differences between the λ estimates of the two different AN subtypes (Fig. 2d). Furthermore, a trend towards a significant correlation ($P = 0.05$) was observed between body mass index and loss aversion parameter λ in the AN group, but not in healthy controls, suggesting that the reduction in loss aversion in AN patients (Fig. 2f) was strongest for those with the lowest body weight.

To test whether the differences in model parameter values between AN patients and controls were sufficient to describe the observed changes in the classic measure of IGT performance (Fig. 2b), we performed a posterior predictive check of the model[34]. To this aim, we simulated data for each participant individually using only the participant's model parameter estimates, and plotted the IGT performance of the simulated data over the different trial blocks. This procedure replicated the observed impairment in IGT performance (Fig. 3), indicating that the differences in model parameter values were sufficient to explain differences in IGT performance in the group.

Replication in second cohort

In order to replicate the observed effects, we tested an additional 216 patients on the IGT; 142 with the restrictive subtype and 74 patients with the binge-purge subtype. Although this experiment lacked a formal healthy control group, assessing the absolute value estimate of loss aversion parameter λ may provide insights into the IGT performance of this patient group. Again, we observed a λ parameter value estimate that was close to 1

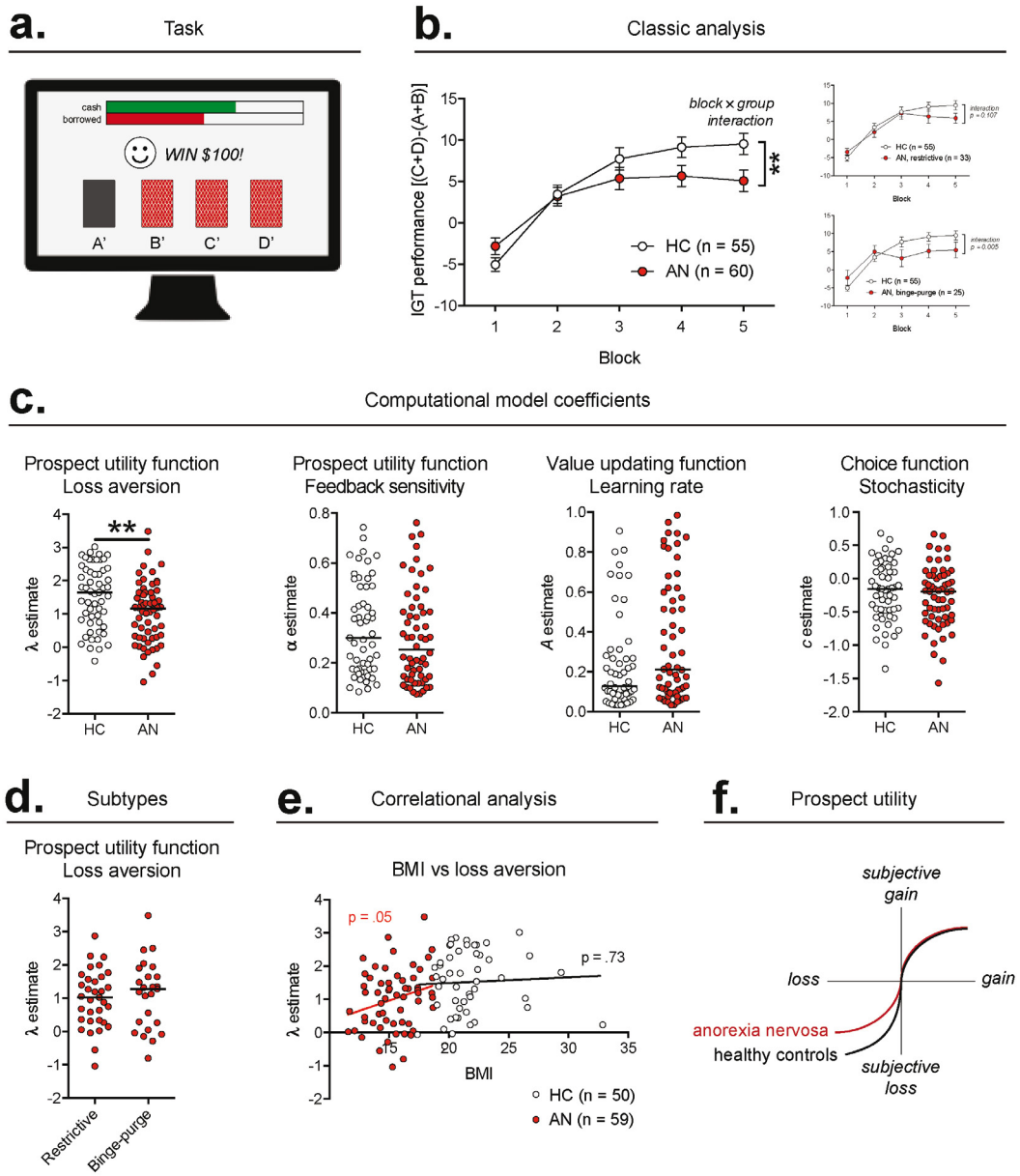


Figure 2: AN patients show reduced IGT performance. **a.** Task design. Losses and gains were accompanied by visual (smileys) and auditory (happy vs sad sounds) feedback. A total of 100 trials were performed per participant. **b.** A significant interaction effect was found in the IGT score over the different 20-trial blocks of AN patients compared to controls (2-way ANOVA, main effect of group, p = 0.163; group x block interaction effect, P = 0.004). Two patients were not subtype-classified. **c.** Computational model analysis revealed that AN patients had a lower value of the loss aversion parameter λ (Mann-Whitney test, P = 0.004), indicating that AN patients are less loss averse than controls. No effects were observed on feedback sensitivity parameter α (Mann-Whitney test, P = 0.2158), learning rate A (Mann-Whitney test, P = 0.0535) and stochasticity factor c (unpaired t-test, P = 0.3515). Horizontal lines denote median value. **d.** No effect between the value of parameter λ between the two subtypes of AN (unpaired t-test, P = 0.6200). Horizontal lines denote median value. **e.** Estimates of loss aversion parameter λ showed a trend towards a positive correlation with BMI in AN patients (P = 0.05, R2 = 0.06), but not in controls (P = 0.73, R2 < 0.01). No BMI data was available for six participants. **f.** Visual summary: AN patients are less sensitive to monetary losses than controls.

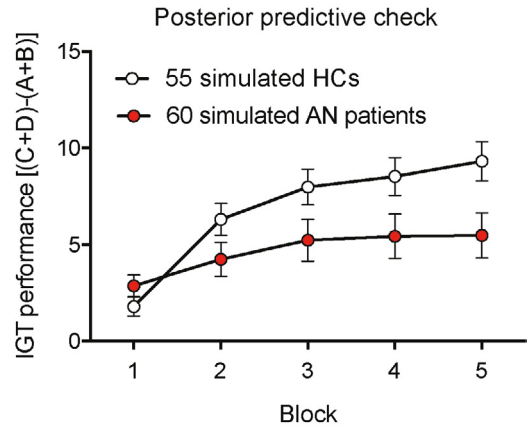


Figure 3: Posterior predictive check of the model. Simulating data with the extracted model parameter estimates (average of 5 simulations per subject) replicated the difference in IGT performance between AN patients and controls (2-way ANOVA, main effect of group, P = 0.0623, group x block interaction effect, P < 0.0001).

(Supplementary Fig. 1a,b); this is considerably lower than what is known in healthy subjects from literature[25,35,36], and significantly lower than the control group from the first cohort (*P* = 0.003), but statistically indistinguishable from the first cohort of AN patients (*P* = 0.40). Furthermore, we again observed no significant differences in the estimates of *λ* between the different AN subtypes (*P* = 0.71). No significant correlation was observed between the estimate of loss aversion parameter *λ* and body mass index (Supplementary Fig. 1c).

Discussion

In this study, we have assessed behavior in the IGT of AN patients and healthy controls, by employing computational trial-by-trial analyses. We replicated data from a vast body of literature[23] that shows that AN patients are impaired in IGT performance, and subsequently demonstrated that this was driven by the absence of loss aversion. Such a decreased sensitivity to monetary losses prevented participants to avoid the disadvantageous decks, leading to a less steep learning curve in the classic measures of IGT performance (Fig. 2b). This diminished loss aversion might also be related to the apparent insensitivity of AN patients to the negative consequences of the disease itself, including the suppression of extreme hunger and social isolation. Interestingly, AN patients self-report *increased* sensitivity to punishment in questionnaires[37], suggesting suboptimal reflection of their own behavior. This mismatch between self-report measures and empirical measures may be of importance, since it sheds light on the ability of AN patients to assess their own actions in hindsight and reflect on their own well-being and body weight.

Several studies have assessed the neural basis of loss aversion in human subjects. In accordance with reward prediction error theory[38,39], one seminal study demonstrated that in healthy subjects that were confronted with different gambling options, fMRI BOLD signal was reduced in the ventral striatum and ventromedial prefrontal cortex, the target regions of midbrain dopamine, when the potential losses of the gamble increased[40]. Given the existence of alterations in the dopamine system in AN patients[17], it is tempting to speculate that malfunction in the dopamine system underlies the observed absence of loss aversion

in AN patients. Indeed, our lab recently showed that negative feedback learning, closely related to adapting to monetary losses, is diminished during an abundance of dopamine in the ventral striatum[16].

Other neuroimaging studies have confirmed alterations in the dopamine system of AN patients. For example, dopamine D2/D3 receptor availability has been shown to be higher in recovered AN patients[18] (but see ref. 41). Furthermore, neural responses to taste reward were increased in midbrain dopamine projection areas in AN patients[19], indicative of alterations in the neural signals associated with value-based learning. However, since the authors did not distinguish between positive and negative prediction errors (guiding reward and punishment learning, respectively), these signals could also represent the general salience of food. Another study compared BOLD responses of recovered AN patients with controls during a monetary reward task, and showed that negative feedback signals in the ventral striatum were largely diminished in AN patients, as this area exhibited similar responses to monetary gains and losses, in contrast to healthy control who showed opposite neural responses to these two types of outcome[42]. Taken together, these studies suggest that reward prediction error signals arising from midbrain dopamine circuits are impaired in AN, which may give rise to disturbances in adapting behavior to feedback, in particular following losses. To date, however, results with psychotropic medication targeting the dopamine system in AN patients have been disappointing, although there is some evidence that treating patients with a dopamine D2 receptor antagonist may have beneficial effects on weight restoration[43]. Besides a role for dopamine, alterations in other brain structures involved in value computations or negative emotion processing could underlie the observed changes in loss aversion in AN patients. For example, structural and neurophysiological changes in the prefrontal cortex and amygdala have been reported in AN patients[44].

Limitations and considerations

One possible concern of this study is the finding that we observed a trend towards a significant correlation between body mass index and the estimate of loss aversion parameter λ in AN patients of the first cohort. Although this may merely reflect the severity and progression of the disease itself, it may also arise from malnutrition in AN patients. Although very few studies have investigated the effects of hunger on performance in standardized decision making tasks, it is generally assumed that a negative energy balance negatively affects cognitive performance[45]. Studies that compare decision making capacity between recovered and symptomatic AN patients demonstrate the persistence of cognitive deficits after recovery[46-48], although the few studies that made this comparison with regards to IGT performance suggest a partial restoration back to normal levels[23].

A limitation of this study is the fact that the patient and control groups of the first cohort differed in terms of their level of education (Table 1). Given the negative effect that AN can have on school performance, however, this does not necessarily imply differences in intelligence. Furthermore, paradoxically, a higher level of education is usually associated with worse IGT performance[49]. Thus, if any effects were to be expected on the basis of education, this effect should be into the opposite direction of what was observed in this study.

Finally, it is interesting to note that the overall estimates of choice stochasticity parameter c were negative for both the control and AN group. This indicates that the majority of participants chose more randomly as the session progressed, which may seem counterintuitive, given that a more exploitative approach would be beneficial once participants have a better understanding of the reward contingencies of the decks later on in the task. Such negative values for this parameter have been found before, also in healthy individuals, and possibly reflect fatigue or boredom[25].

Comparison with other studies

A previous study has also used a computational model based on prospect utility theory to fit a pooled set of IGT data of AN patients, gathered in three independent institutes[24]. Besides the model parameters based on prospect utility theory, their model included the decay-reinforcement learning rule and trial-independent choice rule — a model that we also tested but that was not the best descriptor of IGT performance in our analysis (Table 2, model #8). Furthermore, data from their AN group was collected across three research institutes, while their control group comprised only participants recruited at one institute. Interestingly, they also observed a significant decrease in loss aversion when only comparing patients and controls from the institute that was used for recruitment of the control group, but this effect statistically disappeared after pooling the groups from the different institutes.

A recent study assessed IGT performance in a large cohort of 611 female individuals, approximately half of which were AN patients[50]. Despite their large sample size, they did not observe any significant differences on the model parameters that were fitted to the data, although they used a model based on expected utility theory, that we and others have shown to be a poor descriptor of behavior in the IGT[25]. Furthermore, the authors did not describe what method they used to estimate the model parameters, nor did they provide a quantification of the fit of the model, making a comparison between our studies difficult.

Several other studies have demonstrated altered negative feedback learning in AN patients, although not all of these studies seem directly in accordance with our findings. For example, one study performed a probabilistic reversal learning task in AN patients in an fMRI scanner[51], and observed an increased learning rate for negative feedback in AN patients compared to healthy controls, although this effect was numerically modest. Negative feedback co-incided with elevated levels of activity in the posterior medial prefrontal cortex in AN patients compared to controls, while no differences were observed in hemodynamic responses to reward. Although the dissociable effect on punishment and not on reward is in accordance with our study, we would have expected a decrease, rather than an increase, in learning rate following punishment.

Concluding remarks

Although the neural underpinnings of AN are largely unknown, it has been proposed that the neurocognitive deficits associated with AN are a contributing factor in the progression of the disease and the inability of patients to recover. By using computational analysis of data of patients performing the IGT, we show that AN patients do not exhibit loss aversive behavior, in contrast to what is seen in healthy controls. Our data are in line with previous work that shows alterations in negative feedback processing in AN patients, and points towards disruptions in the brain circuits involved in value processing. Together, these findings provide possible handles for the psychological treatment of AN patients, for the development of pharmacological therapies for AN and provide important fundamental insights in the etiology of the disease.

References

1. Szmukler, G. I. et al. Neuropsychological impairment in anorexia nervosa: before and after refeeding. Journal of Clinical and experimental Neuropsychology 14, 347-352 (1992).
2. Cavedini, P. et al. Neuropsychological investigation of decision-making in anorexia nervosa. Psychiatry research 127, 259-266 (2004).
3. Cavedini, P. et al. Decision-making functioning as a predictor of treatment outcome in anorexia nervosa. Psychiatry research 145, 179-187 (2006).
4. Kaye, W. H., Fudge, J. L. & Paulus, M. New insights into symptoms and neurocircuit function of anorexia nervosa. Nature Reviews Neuroscience 10, 573 (2009).
5. Brogan, A., Hevey, D. & Pignatti, R. Anorexia, bulimia, and obesity: shared decision

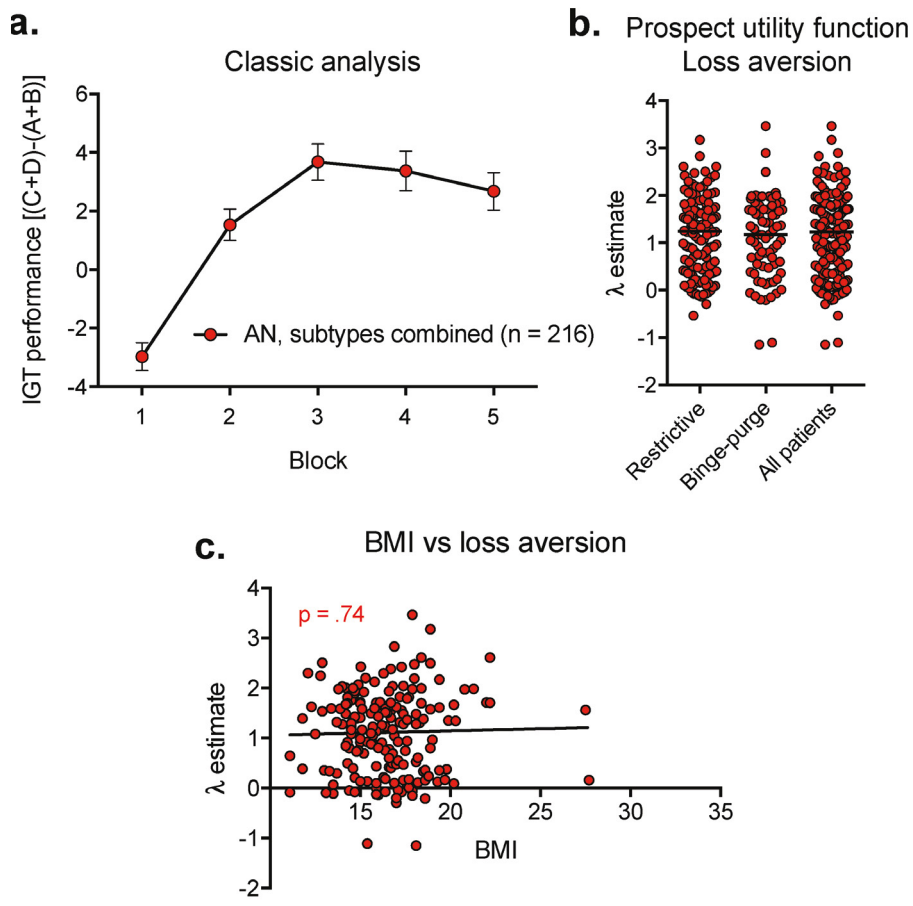
making deficits on the Iowa Gambling Task (IGT). Journal of the International Neuropsychological Society 16, 711-715 (2010).
6. Danner, U. N. et al. Neuropsychological weaknesses in anorexia nervosa: Set shifting, central coherence, and decision making in currently ill and recovered women. International Journal of Eating Disorders 45, 685-694 (2012).
7. Lindner, S. E., Fichter, M. M. & Quadflieg, N. Decision making and planning in full recovery of anorexia nervosa. International Journal of Eating Disorders 45, 866-875 (2012).
8. Tchanturia, K. et al. Poor decision making in male patients with anorexia nervosa. European Eating Disorders Review 20, 169-173 (2012).
9. Friederich, H. & Herzog, W. in Behavioral neurobiology of eating disorders 111-123 (Springer, 2010).
10. Steinglass, J. E., Walsh, B. T. & Stern, Y. Set shifting deficit in anorexia nervosa. Journal of the International Neuropsychological Society 12, 431-435 (2006).
11. Roberts, M. E., Tchanturia, K., Stahl, D., Southgate, L. & Treasure, J. A systematic review and meta-analysis of set-shifting ability in eating disorders. Psychological medicine 37, 1075-1084 (2007).
12. Steinglass, J. E. et al. Increased capacity to delay reward in anorexia nervosa. Journal of the International Neuropsychological Society 18, 773-780 (2012).
13. Lauer, C. J., Gorzewski, B., Gerlinghoff, M., Backmund, H. & Zihl, J. Neuropsychological assessments before and aftertreatment in patients with anorexia nervosa and bulimianervosa. Journal of Psychiatric Research 33, 129-138 (1999).
14. Smith, K. E., Mason, T. B., Johnson, J. S., Lavender, J. M. & Wonderlich, S. A. A systematic review of reviews of neurocognitive functioning in eating disorders: The state-of-the-literature and future directions. International Journal of Eating Disorders (In press), doi:10.1002/eat.22929.
15. Rangel, A., Camerer, C. & Montague, P. R. A framework for studying the neurobiology of value-based decision making. Nature reviews neuroscience 9, 545 (2008).
16. Verharen, J. P. H. et al. A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. Nature communications 9 (2018).
17. Kaye, W. H., Frank, G. K. W. & McConaha, C. Altered dopamine activity after recovery from restricting-type anorexia nervosa. Neuropsychopharmacology 21, 503 (1999).
18. Frank, G. K. et al. Increased dopamine D2/D3 receptor binding after recovery from anorexia nervosa measured by positron emission tomography and [11c] raclopride. Biological psychiatry 58, 908-912 (2005).
19. Frank, G. K. W. et al. Association of Brain Reward Learning Response With Harm Avoidance, Weight Gain, and Hypothalamic Effective Connectivity in Adolescent Anorexia Nervosa. JAMA Psychiatry (2018).
20. Bechara, A., Damasio, A. R., Damasio, H. & Anderson, S. W. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 50 (1994).
21. Bechara, A., Tranel, D., Damasio, H. & Damasio, A. R. Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. Cerebral Cortex 6 (1996).
22. Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. Deciding advantageously before knowing the advantageous strategy. Science 275 (1997).
23. Guillaume, S. et al. Impaired decision-making in symptomatic anorexia and bulimia nervosa patients: a meta-analysis. Psychol Med 45, 3377-3391 (2015).
24. Chan, T. W. et al. Differential impairments underlying decision making in anorexia nervosa and bulimia nervosa: a cognitive modeling analysis. Int J Eat Disord 47, 157-167 (2014).
25. Ahn, W. Y., Busemeyer, J. R., Wagenmakers, E. J. & Stout, J. C. Comparison of decision learning models using the generalization criterion method. Cogn Sci 32, 1376-1402 (2008).
26. Kahneman, D. Prospect theory: An analysis of decisions under risk. Econometrica 47, 278 (1979).
27. Tversky, A. & Kahneman, D. The framing of decisions and the psychology of choice. science 211, 453-458 (1981).
28. Harrison, G. W. & Rutström, E. E. Expected utility theory and prospect theory: One wedding and a decent funeral. Experimental Economics 12, 133 (2009).
29. Steingroever, H., Wetzels, R. & Wagenmakers, E. J. Validating the PVL-Delta model for the Iowa gambling task. Front Psychol 4, 898 (2013).
30. Yechiam, E. & Busemeyer, J. R. Comparison of basic assumptions embedded in learning models for experience-based decision making. Psychonomic Bulleting & Review 12, 387-402 (2005).
31. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies - revisited. Neuroimage 84, 971-985 (2014).
32. Tversky, A. & Kahneman, D. Loss aversion in riskless choice: A reference-dependent model. The quarterly journal of economics 106, 1039-1061 (1991).
33. Trepel, C., Fox, C. R. & Poldrack, R. A. Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. Cognitive brain research 23, 34-50 (2005).
34. Gelman, A., Meng, X. & Stern, H. Posterior predictive assessment of model fitness via realized discrepancies. Statistica sinica, 733-760 (1996).
35. Vassileva, J. et al. Computational modeling reveals distinct effects of HIV and history of drug use on decision-making processes in women. PLoS One 8, e68962 (2013).
36. Worthy, D. A., Hawthorne, M. J. & Otto, A. R. Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. Psychonomic Bulletin & Review 20, 364-371 (2013).
37. Jappe, L. M. et al. Heightened sensitivity to reward and punishment in anorexia nervosa. Int J Eat Disord 44, 317-324 (2011).
38. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. Science 275 (1997).
39. Schultz, W. Dopamine reward prediction-error signalling: a two-component response. Nature reviews. Neuroscience 17, 183-195 (2016).
40. Tom, S. M., Fox, C. R., Trepel, C. & Poldrack, R. A. The Neural Basis of Loss Aversion in Decision-Making Under Risk. Science, 515-518 (2007).
41. Broft, A. et al. Striatal dopamine type 2 receptor availability in anorexia nervosa. Psychiatry Research: Neuroimaging 233, 380-387 (2015).
42. W., A. et al. Altered Reward Processing in Women Recovered From Anorexia Nervosa. American Journal of Psychiatry 164, 1842-1849 (2007).
43. Frank, G. K. & Shott, M. E. The role of psychotropic medications in the management of anorexia nervosa: rationale, evidence and future prospects. CNS drugs 30, 419-442 (2016).
44. Titova, O. E., Hjorth, O. C., Schiöth, H. B. & Brooks, S. J. Anorexia nervosa is linked to reduced brain structure in reward and somatosensory regions: a meta-analysis of VBM studies. BMC Psychiatry 13, 110 (2013).
45. Keys, A., Brožek, J., Henschel, A., Mickelsen, O. & Taylor, H. L. The biology of human starvation. (1950).
46. Bosanac, P. et al. Neuropsychological study of underweight and “weight recovered” anorexia nervosa compared with bulimia nervosa and normal controls. International Journal of Eating Disorders 40, 613-621 (2007).
47. Cowdrey, F. A., Park, R. J., Harmer, C. J. & McCabe, C. Increased neural processing

of rewarding and aversive food stimuli in recovered anorexia nervosa. Biological psychiatry 70, 736-743 (2011).
48. Frank, G. K., Shott, M. E., Hagman, J. O. & Mittal, V. A. Alterations in brain structures related to taste reward circuitry in ill and recovered anorexia nervosa and in bulimia nervosa. American Journal of Psychiatry 170, 1152-1160 (2013).
49. Evans, C. E. Y., Kemish, K. & Turnbull, O. H. Paradoxical effects of education on the Iowa Gambling Task. Brain and Cognition 54, 240-244 (2004).
50. Giannunzio, V. et al. Decision-making impairment in anorexia nervosa: New insights into the role of age and decision-making style. European Eating Disorders Review 26, 302-314 (2018).
51. Bernardoni, F. et al. Altered Medial Frontal Feedback Learning Signals in Anorexia Nervosa. Biol Psychiatry 83, 235-243 (2018).

Conflict of interest
The authors declare no conflict of interest.

Acknowledgements
J.P.H.V. was funded by the European Union Seventh Framework Programme under grant agreement number 607310 (*Nudge-IT*).

SUPPLEMENTARY FIGURE 1



Data second cohort
a. IGT performance across trial blocks in the second cohort of AN patients. This learning curve is not significantly different from the first cohort of AN patients (two-way ANOVA, main effect of group, P = 0.0640; group x block interaction effect, P = 0.5773) but it is significantly different from the control group of the first cohort (two-way ANOVA, main effect of group, P = 0.0003; group x block interaction effect, P < 0.0001).
b. Value estimates of loss aversion parameter λ of the second cohort of AN patients. These values were not significantly different from the λ estimates of the first cohort of AN patients (unpaired t-test, all patients second cohorts versus all patients first cohort, P = 0.3955), but they were significantly different from the control group of the first cohort (Mann-Whitney test, P = 0.0028). No significant differences were found between the restrictive and binge-purge subtype of AN (Mann-Whitney test, P = 0.7147). Horizontal lines denote median value.
c. In the second cohort, the estimate of loss aversion parameter λ did not significantly correlate with body mass index (R2 < 0.01, p = 0.74). Based on n = 185 participants; body mass index information was not available for 31 participants.

CHAPTER 10

General discussion

CHAPTER 10

In this thesis, I sought to gain insights into the neurocomputational basis of decision making and motivation, by looking at how reward and aversion shape choice behavior of rats (chapters 2 through 8) and humans (chapter 9). By combining different behavioral, pharmacological, genetic and computational tools, we were able to increase our understanding of the neural mechanisms involved in adapting to positive and negative feedback and the process in which the costs and benefit of decisions are weighed. These findings provide important fundamental insights into the neurobiology of decision making and motivation, which can help understand the pathophysiology of mental disorders associated with deficits in these processes.

The neuronal basis of value-based learning and decision making

In **chapter 2** of this thesis (see also **Box 1**), we studied the effects of chemogenetic stimulation of the two major dopaminergic pathways of the mesocorticolimbic system: the ventral tegmental area-to-nucleus accumbens pathway (VTA → NAc) and the VTA-to-medial prefrontal cortex pathway (VTA → mPFC). In one of the first behavioral experiments that we conducted, we observed that stimulation of Gq-coupled designer receptors on VTA → NAc neurons made rats insensitive to punishment, in that once they started lever pressing for sucrose, they did not suppress this behavior if these presses were followed by electric foot shock punishment. After also observing that stimulation of this pathway impaired adaptation to negative feedback (a reward omission or 'loss') in a serial reversal learning task, we hypothesized that this behavior may be related to impaired processing of negative reward prediction error signals in the NAc, as was previously hypothesized to be involved in the etiology of dopamine dysregulation syndrome in Parkinson's disease[1-3], a well-known condition associated with an abundance of dopamine in the brain[4,5]. After conducting a large array of behavioral experiments which we combined with chemogenetics, pharmacology, fiber photometry and microdialysis, we concluded that hyperactivity of the VTA → NAc pathway indeed evoked a phenotype of loss and punishment insensitivity, and that this is likely related to 'overdosing' dopamine receptors in the NAc with dopamine, which makes the NAc unable to detect the transients dips in dopamine release that are associated with negative feedback. Besides the possible involvement of this mechanism in the etiology of dopamine dysregulation syndrome, we speculated that this may also be involved in the overoptimistic and reckless behaviors observed during the 'high' of drugs and during the manic phase of bipolar disorder, two other conditions that are associated with increased extracellular concentrations of dopamine[6-9].

For **chapters 3 through 5**, we modified the serial reversal learning task that we had used in chapter 2, to make it more suitable for computational modeling analysis by rendering the reward contingencies probabilistic. In this new version of the task, a lever press at the 'active', or high-probability lever (or nosepoke hole) resulted in reinforcement in 80% (rather than 100%) of trials, while pressing the low-probability lever was reinforced in 20% (rather than 0%) of trials, so that the animals also had to process false information with regards to which of the two options is most beneficial. In this way, animals had to track the value of the two choice options by integrating a longer history of outcomes and make a choice based hereon. We fit different computational reinforcement learning models to datasets of rats performing this probabilistic reversal learning task, and found that a model based on the classic Rescorla-Wagner learning theory[10,11] predicted task behavior best, both in male (model comparison performed in chapter 3) and female rats (chapter 5). This model assesses trial-by-trial data of the rats and describes behavior of the animals on the basis of four parameters (Figure 1): the reward learning rate, describing the extent to which a single reinforced trial increases lever value; the punishment learning rate, describing to what extent a single non-reinforced trial decreases lever value; a stickiness parameter that describes the amount of perseveration of responding on the same lever; and a stochasticity parameter that describes the number of explorative choices (i.e., choosing the lowest valued lever) compared to the number of exploitative choices (i.e., choosing the highest valued lever).

In **chapter 3**, we fit this model to the behavioral data of rats after pharmacological inactivation of different regions of the prefrontal cortex by infusion of a cocktail of the GABA receptor agonists baclofen and muscimol. We found that after inactivation of any of the four studied prefrontal cortex regions (the prelimbic, infralimbic, medial orbitofrontal and lateral orbitofrontal cortices), the value estimate of the punishment learning rate was decreased, indicating a lower impact of negative feedback on behavior. Although a decreased learning rate does not imply an impairment *per se* (since for probabilistic reversal learning a wider range of learning rates can lead to successful task execution), it does suggest an involvement of these four brain regions in negative feedback processing. Similarly, we found an involvement of the prelimbic and lateral orbitofrontal cortex in positive feedback learning and of the infralimbic and medial orbitofrontal cortex in choice perseveration. Besides these alterations in the value estimates of the model parameters, we also observed changes in the classic measures of task performance; inactivation of the infralimbic and lateral orbitofrontal cortex reduced the total number of reversals, and inactivation of the prelimbic and medial orbitofrontal cortex reduced the total number of rewards the animals obtained. Together, these data shed light on the complexity of value-based learning and decision making, and provide evidence that value is processed in multiple brain circuits in parallel, as has been suggested by recent theories[12-14]. It further shows robust, whole-prefrontal cortex coding of negative feedback learning, which may be related to the importance of this type of learning to survival — after all, after experiencing danger, an organism must learn not to get caught up in that same situation in the future.

In **chapter 4**, this same computational approach provided insight into a long-standing question in behavioral pharmacology: how dopamine D1- and D2-receptor expressing neurons in the striatum contribute to value-based learning and decision making. Influential theories have suggested that these two types of neurons have an opposing function in approach versus avoidance learning, respectively[15-17]. Our data support this theory by showing that local infusion of D1 receptor agonist SKF82958 into the ventral striatum affected positive feedback learning while infusion of D2 receptor agonist quinpirole affected negative feedback learning. We found an additional function of the dopamine D2 receptor in the ventral and dorsolateral, but not dorsomedial, striatum in mediating the exploration/exploitation balance, a parameter related to the decision-making aspect of motivated behavior (Figure 1). The findings from this chapter also shed light on the findings of chapter 2, in which we found that a hyperdopaminergic state is associated with attenuated negative feedback learning, in that these effects are probably mediated through dopamine D2 receptor activation in the ventral striatum.

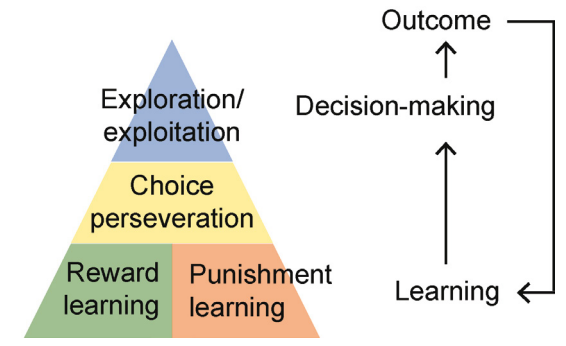


Figure 1: parameters of computational model

In **chapter 5**, we studied whether the four assessed parameters of value-based decision making fluctuate across the estrous cycle of female rats. We found that positive feedback learning and the exploration/exploitation balance (as well as motivation to obtain food reward) fluctuated across the cycle, while negative feedback learning and choice perseveration did not. We speculated that these fluctuations promote adaptive survival-directed behavior during certain stages of the cycle through action of gonadal steroids in the mesocorticolimbic system[18]. Through which neuronal mechanisms these effects are mediated and why this is evolutionary beneficial are questions of outstanding interest, answering of which requires further study.

Motivational aspects of decision making

Many decisions we make in everyday life are complex by nature, and require a careful balancing of the costs and benefits associated with different courses of action. For example, during food choices of humans, the taste and content of a certain food are considered in combination with its health consequences and costs of purchase. As a result, some individuals are better in making healthy food choices than others, probably because they weigh these different factors differently in their decision-making process, or because baseline levels of health and wealth are different. A special case in which food-related decision making goes ostensibly awry is during binge eating, in which (usually unhealthy) foods are consumed in amounts higher than initially planned and desired, i.e., people lose control over food intake.

In an attempt to model this behavior in rats, we set up a novel behavioral task in **chapter 6** that studies the animals' ability to inhibit the urge to consume a tasty sucrose pellet. In this task, animals received a pellet at the start of every 40-second trial, which they could consume freely in half of the trials, but needed to wait with consumption of the pellet during the other half of the trials. In these waiting trials, behavioral control was contingent on the presence of an audiovisual stimulus that acted as a danger signal for the animals, and animals quickly learned to control their behavior during this cue. Benefits of this task, in comparison with comparable tasks, are that it requires relatively little training, it is able to discern between different behavioral phenotypes (including loss of behavioral control, loss of stimulus retrieval and loss of motivation), and it tests control over the intake of a primary reinforcer, which may provide a more naturalistic approach to behavioral control than tasks that involve responding on arbitrary manipulanda, such as levers or nosepoke holes. As such, pharmacological inactivation of any of the regions in the ventromedial prefrontal cortex (infralimbic, prelimbic or medial orbitofrontal cortex) evoked loss of control over behavior in the animals, so that they repeatedly reached out for the food pellet during the presentation of the audiovisual stimulus, despite getting foot shock punishment. Inactivation of the basolateral amygdala also evoked a phenotype in which animals continuously reached out for the pellet despite the presence of the threat signal, although additional latency analyses suggested that this was driven by a role of the basolateral amygdala in retrieving the value of the stimulus, rather than by directly mediating behavioral control. By combining this task with chemogenetics, fiber photometry and pharmacology in **chapter 7**, we found surprisingly little evidence for a direct involvement of dopaminergic neurotransmission in behavioral control.

In **chapter 8**, we studied the contribution of the mesocorticolimbic system to sodium appetite, which is a special case of a costs/benefit decision. Sodium appetite refers to the fact that a sodium deficiency of the body evokes vigorous cravings for salty foods, while salt is normally considered aversive by many organisms[19-22]. Salt can therefore act as a cost (i.e., a punisher) or as a benefit (i.e., a reward), depending on the physiological state of the animal. The mechanism by which the brain can make this 'switch' in salt appreciation from aversive to appetitive remains elusive, and studies about a direct involvement of the dopamine system in salt appetite have been inconclusive (e.g., see refs. 23-26). Here, we tried to resolve these inconsistencies by using fiber photometry, c-Fos immunoreactivity, chemogenetics

Box 1

The most important conclusions from the chapters in one sentence.

Chapter 2

- An abundance of dopamine in the brain evokes a phenotype of loss and punishment insensitivity through overstimulation of dopamine receptors in the nucleus accumbens and the subsequent impairment in negative prediction error processing.

Chapter 3

- Punishment learning is dependent on a wide array of prefrontal cortex regions (prelimbic, infralimbic and orbitofrontal cortices), while reward learning and choice perseveration are anatomically segregated.

Chapter 4

- Stimulation of dopamine D1 and D2 receptors in the ventral striatum guides reward and punishment learning, respectively, while exploratory choice behavior is dependent on the dopamine D2 receptor in the ventral and dorsolateral striatum.

Chapter 5

- Reward learning, exploration and motivation fluctuate across the estrous cycle of female rats.

Chapter 6

- Using a newly developed behavioral task for rats, we show that the medial prefrontal cortex is important for inhibitory control over behavior.

Chapter 7

- There is little evidence of a direct involvement of dopaminergic neurotransmission in the exertion of behavioral control in rats.

Chapter 8

- The nucleus accumbens may mediate the motivational, but not hedonic, component of sodium appetite in rats, but this is not driven by dopamine.

Chapter 9

- Anorexia nervosa patients show insensitivity to monetary losses in the Iowa gambling task in comparison to healthy controls.

and behavioral pharmacology to study the involvement of the mesocorticolimbic dopamine system in sodium appetite. By using a microstructural analysis of licking behavior[27-30] during a home cage free-intake paradigm, we tried to parse the effects of dopaminergic manipulations on the motivational versus hedonic component of sodium appetite. We show that its motivational, but not hedonic, aspect is likely dependent on functional activity in the ventral striatum, but that this effect is not mediated by dopaminergic neurotransmission in this area.

Value-based decision making in neuropsychiatric conditions

As reviewed in the introduction of this thesis, abnormalities in the brain circuits involved in value-based decision making and motivation have been implicated in a wide range of neuropsychiatric conditions. One of these conditions is the eating disorder anorexia nervosa, which has been associated with alterations in dopamine neurotransmission in the mesocorticolimbic system[31,32]. Furthermore, anorexia nervosa patients show impairments in laboratory tasks that assess decision making capacity, such as set shifting[33,34] and the Iowa gambling task[35]. In an attempt to find the computational basis of these decision making deficits, and to bridge the gap between decision making tasks in rodents and men, we used computational modeling of data of a group of anorexia nervosa patients and controls that performed the Iowa gambling task in **chapter 9**. We showed that a model based on prospect utility theory (which assumes that the subjective pleasure from a monetary win is not linearly proportional to its numerical size) is superior in explaining the participants' choices, and subsequently showed that anorexia nervosa patients are less sensitive to monetary losses than healthy controls. We speculated that this loss insensitivity is related to the disturbances in the dopamine system that have been observed in anorexia nervosa patients, and thereby provide a possible framework for the pharmacological treatment of these patients. Interestingly, certain studies have suggested an increased dopaminergic tone in anorexia nervosa patients, and we showed, in chapter 2, that an abundance of dopamine in the brain, at least in rats, leads to insensitivity to negative feedback. Although it is tempting to speculate that these changes in value-based decision making are the direct result of increased dopaminergic tone in anorexia nervosa patients, further research should decipher whether this is indeed the case.

Methodological considerations

Most experimental work in this thesis has been performed in rats. Although the brains of humans and rodents show structural and functional similarities, it remains a question whether our findings are directly applicable to humans. In this regard, it is especially challenging to create behavioral tasks that are unambiguous in their interpretation, and of which the outcome parameters demonstrate face, predictive and construct validity. Of course, decision making behavior of humans is a lot more complex by nature than the decisions observed in species like mice and rats. For example, the long-term negative health consequences of certain actions, like taking unhealthy foods or using drugs, are simply not a factor when animals make choices. It is therefore very difficult, if not impossible, to develop behavioral tasks that encompass bad eating habits or addiction as a whole. It is, however, possible to model certain aspects of these phenomena, like sensitivity to reward or simple economic considerations (such as choosing between a small reward now or a large reward in a minute from now).

Computational modeling of behavioral data was one way for us to increase the translational value of our animal models, thereby assessing subtle changes in the strategy of animals in a reversal learning paradigm. The parameters that can be extracted from such a computational model provide knowledge about the processes that comprise the basis of complex decision making behavior, such as learning from reward and punishment, and perseverative behavior. In fact, these same computational models can be applied to human data of tasks like the two-armed bandit task[36,37], which is essentially a version of probabilistic reversal learning. If this is done thoroughly, data from rodent studies can be used to directly test theories from computational psychiatry that have been based on human studies, thereby utilizing the extensive toolbox of neural manipulations that is available for rodents.

The novel task that models control over behavior that we presented in chapters 6 and 7 is an attempt to provide a more naturalistic approach to behavioral control. Many tasks that study similar behaviors, such as the 5-choice serial reaction time task and the stop-signal task, study behavioral control based on responses on arbitrary manipulanda, like nosepokes or lever presses. The fact that in our task, rats had to control themselves at the mere sight of a food reward can therefore be seen as better reflective of the human situation, in which food is abundantly available but must be consumed in limited amounts in order to stay healthy. That said, the punishment aspect of behavioral tasks, like in this case an electric foot shock, have limited translatable value, since the punishment involved in human decision-making processes are often more probabilistic and long-term by nature, such as the negative health consequences that develop over a longer period of time.

Future directions

"*Science is always wrong. It never solves a problem without creating 10 more.*"

Stuart Firestein, quoting George Bernard Shaw in his
book Ignorance: How It Drive Science (2012)

In this thesis, I have assessed how the brain of the rat processes reward and aversion and how that eventually leads to adaptations in behavior. We have, for example, learned that the neural mechanisms of reward and punishment learning are partially segregated in the forebrain, that deviations from an optimum in neurotransmitter concentrations can hamper the computations associated with decision making, that dopamine is not involved in all motivational processes, and that some disease states are associated with changes in value-based decision making. As such, I have answered some important fundamental questions about the basic processes underlying decision-making and motivation. That said, as set forth by Stuart Firestein in his book '*Ignorance: How It Drives Science*', scientific experiments should raise more questions than they answer. Therefore, in **box 2**, I conclude this thesis with an overview of the most important and most pressing questions that arise from each of the eight experimental chapters.

Box 2

Questions of outstanding interest

- To what extent is the overoptimistic behavior seen during hyperdopaminergic states the result of impaired negative feedback learning, as compared to other cognitive effects of an abundance of dopamine in the brain, such as enhanced temporal discounting capacity?
- Why is punishment learning so abundantly coded in the prefrontal cortex?
- What is the computational process that underlies explorative versus exploitative choice behavior, and why is this only dependent on the dopamine D2 receptor?
- What is the evolutionary advantage of adaptive changes in reinforcement learning during the different stages of the rat estrous cycle?
- Through which motor pathway does the medial prefrontal cortex exert control over behavior?
- Why are monoamine reuptake inhibitors effective for certain impulse control disorders, if the dopamine system does not directly mediate behavioral control?
- Why are some forms of motivation dependent on dopaminergic neurotransmission in the nucleus accumbens, and others not?
- Are the differences in value processing in anorexia nervosa patients involved in the etiology of anorexia nervosa or are they merely an epiphenomenon of the disease?

References

1. Cools, R., Barker, R. A., Sahakian, B. J. & Robbins, T. W. Enhanced or Impaired Cognitive Function in Parkinson's Disease as a Function of Dopaminergic Medication and Task Demands. Cereb Cortex 11, 1136-1143 (2001).
2. Frank, M. J., Seeberger, L. C. & O'Reilly R, C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306, 1940-1943 (2004).
3. Frank, M. J. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. Journal of Cogn Neurosci 17, 51-72 (2005).
4. Evans, A. H. & Lees, A. J. Dopamine dysregulation syndrome in Parkinson's disease. Curr Opin in Neurol 17, 393-398 (2004).
5. Berk, M. et al. Dopamine dysregulation syndrome: implications for a dopamine hypothesis of bipolar disorder. Acta Psychiatrica Scandinavica 116, 41-49 (2007).
6. Janowsky, D. S., Davis, J. M., Khaled El-Yousef, M. & Sekerke, H. J. A Cholinergic-Adrenergic Hypothesis of Mania and Depression. The Lancet 2, 632-636 (1972).
7. Di Chiara, G. & Imperato, A. Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. Proc Natl Acad Sci 85, 5274-5278 (1988).
8. Johnson, S. L. Mania and dysregulation in goal pursuit: a review. Clin Psychol Rev 25, 241-262 (2005).
9. Lüscher, C. & Ungless, M. A. The mechanistic classification of addictive drugs. PLoS Med 3, e437 (2006).
10. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 2, 64-99 (1972).
11. Sutton, R. S. & Barto, A. G. Reinforcement learning: An introduction. (MIT press, 1998).
12. Cisek, P. Making decisions through a distributed consensus. Current opinion in neurobiology 22, 927-936 (2012).
13. Rushworth, M. F., Kolling, N., Sallet, J. & Mars, R. B. Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Current opinion in neurobiology 22, 946-955 (2012).
14. Hunt, L. T. & Hayden, B. Y. A distributed, hierarchical and recurrent framework for reward-based choice. Nature Reviews Neuroscience 18, 172 (2017).
15. Surmeier, D. J., Ding, J., Day, M., Wang, Z. & Shen, W. D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. Trends Neurosci 30, 228-235 (2007).
16. Collins, A. G. E. & Frank, M. J. Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychological review 121, 337 (2014).
17. Francis, T. C. & Lobo, M. K. Emerging Role for Nucleus Accumbens Medium Spiny Neuron Subtypes in Depression. Biol Psychiatry 81, 645-653 (2017).
18. McEwen, B. S. & Alves, S. E. Estrogen Actions in the Central Nervous System. Endocrine Reviews 20, 279-307 (1999).
19. Richter, C. Salt appetite of mammals: its dependence on instinct and metabolism. L'Instinct dans le comportement des animaux et de l'homme 577 (1956).
20. Berridge, K. C., Flynn, F. W., Schulkin, J. & Grill, H. J. Sodium Depletion Enhances Salt Palatability in Rats. Beh Neurosci 98, 632-660 (1984).
21. Berridge, K. C. Palatability Shift of a Salt-Associated Incentive During Sodium Depletion. The Quarterly J of Exp Psychol 41B, 121-138 (1989).
22. Roitman, M. F., Schafe, G. E., Thiele, T. E. & Bernstein, I. L. Dopamine and Sodium Appetite: Antagonists Suppress Sham Drinking of NaCl Solutions in the Rat. Beh Neurosci 111, 606-611 (1997).
23. Wolf, G. Hypothalamic regulation of sodium intake: relations to preoptic and tegmental function. Am J of Psychol 213, 1433-1438 (1967).
24. Stricker, E. M. & Zigmond, M. J. Effects on homeostatis of intraventricular injections of 6-hydroxydopamine in rats. J of Comp and Physiol Psych 86, 973-994 (1974).
25. Fortin, S. M. & Roitman, M. F. Challenges to body fluid homeostasis differentially recruit phasic dopamine signaling in a taste-selective manner. The Journal of neuroscience (2018).
26. Sandhu, E. C. et al. Phasic Stimulation of Midbrain Dopamine Neuron Activity Reduces Salt Consumption. eNeuro 5 (2018).
27. Davis, J. D. The Microstructure of Ingestive Behavior. Annals of the New York Academy of Sciences 575, 106-119 (1989).
28. Higgs, S. & Cooper, S. J. Evidence for early opioid modulation of licking responses to sucrose and Intralipid: a microstructural analysis in the rat. Psychopharmacol. 139, 342-355 (1998).
29. D'Aquila, P. S. Dopamine on D2-like receptors "reboosts" dopamine D1-like receptor-mediated behavioural activation in rats licking for sucrose. Neuropharmacology 58, 1085-1096 (2010).
30. Ostlund, S. B., Kosheleff, A., Maidment, N. T. & Murphy, N. P. Decreased consumption of sweet fluids in mu opioid receptor knockout mice: a microstructural analysis of licking behavior. Psychopharmacology (Berl) 229, 105-113 (2013).
31. Kaye, W. H., Frank, G. K. W. & McConaha, C. Altered dopamine activity after recovery from restricting-type anorexia nervosa. Neuropsychopharmacology 21, 503 (1999).
32. Frank, G. K. et al. Increased dopamine D2/D3 receptor binding after recovery from anorexia nervosa measured by positron emission tomography and [11c] raclopride. Biological psychiatry 58, 908-912 (2005).
33. Steinglass, J. E., Walsh, B. T. & Stern, Y. Set shifting deficit in anorexia nervosa. Journal of the International Neuropsychological Society 12, 431-435 (2006).
34. Roberts, M. E., Tchanturia, K., Stahl, D., Southgate, L. & Treasure, J. A systematic review and meta-analysis of set-shifting ability in eating disorders. Psychological medicine 37, 1075-1084 (2007).
35. Guillaume, S. et al. Impaired decision-making in symptomatic anorexia and bulimia nervosa patients: a meta-analysis. Psychol Med 45, 3377-3391 (2015).
36. Sutton, R. S. Temporal credit assignment in reinforcement learning. (1984).
37. Vermorel, J. & Mohri, M. Multi-armed Bandit Algorithms and Empirical Evaluation. In European conference on machine learning. 437-448 (2005).

CHAPTER 11

Addendum

CHAPTER 11

SAMENVATTING IN HET NEDERLANDS

Neuroeconomische mechanismen van beloning en straf

In dit proefschrift heb ik onderzocht hoe het brein beloning en straf verwerkt en hoe dit uiteindelijk leidt tot aanpassingen in gedrag. Door gebruik te maken van de rat als model voor beslisgedrag konden we technieken gebruiken die niet mogelijk zijn bij mensen. Deze technieken zijn onder andere het direct aflezen van activiteit van hersencellen ('fiber photometry'), het tijdelijk stilleggen van bepaalde structuren in het brein om zo te kijken hoe beslisgedrag van ratten verandert ('behavioral pharmacology') en het activeren van bepaalde groepen hersencellen met behulp van door virus ingebrachte eiwitten ('chemogenetics'). Veel van de bevindingen in dit proefschrift hebben betrekking op de prefrontale hersenschors, welke in de mens gelokaliseerd is net boven de ogen, en op de signaalstof dopamine, welke wordt afgegeven door hersencellen in de middenhersenen en die betrokken is bij het leren van beloning en straf.

Een aantal van de belangrijkste bevindingen uit dit proefschrift:

- In hoofdstuk 2 van dit proefschrift laten we zien dat het overoptimistisch en roekeloos gedrag dat we zien als er te veel dopamine in het brein vrijkomt (bijvoorbeeld na drugsgebruik of tijdens de manische fase van bipolaire stoornis) mogelijk wordt veroorzaakt door het onvermogen om te leren van straf. Dit onvermogen wordt veroorzaakt doordat een hersengebied voorin het brein, de 'nucleus accumbens', wordt overspoeld met dopamine en hierdoor negatieve leersignalen niet meer kan verwerken.
- In hoofdstuk 3 inactiveerden we verschillende gebieden van de prefrontale hersenschors in de rat en hebben we onderzocht hoe dit hun beslisgedrag beïnvloedde. We laten zien dat het leren van straf afhankelijk is van grote delen van de prefrontale hersenschors, maar dat andere processen, zoals leren van beloning of repetitief keuzegedrag, door meer nauwkeurig omschreven delen van de prefrontale hersenschors worden gemedieerd.
- In hoofdstuk 6 en 7 hebben we een nieuwe gedragstaak voor ratten ontwikkeld, waarmee we kunnen onderzoeken hoe het brein controle over gedrag uitoefent. Eerst hebben we ratten geleerd om een voedselbeloning die recht voor hen ligt niet te eten wanneer gelijktijdig een gevaarsignaal werd gepresenteerd (bestaande uit een toon en lampje), en vervolgens hebben we activiteit van hersencellen gemanipuleerd om te kijken hoe dit hun vermogen om controle over gedrag uit te oefenen aantast. We laten zien dat de signaalstof dopamine een kleinere rol speelt in dit proces dan eerder gedacht, maar dat het middelste deel van de prefrontale hersenschors belangrijk is voor het onder controle houden van de neiging om voedsel meteen op te eten.
- Hoofdstuk 9 bevat een onderzoek bij mensen, waarin we laten zien dat het verwerken van beloningssignalen verstoord is bij patiënten met anorexia nervosa. Specifiek lijken zij niet gevoelig voor het verliezen van geld in een goktaak.

De bevindingen uit dit proefschrift zijn voornamelijk fundamenteel van aard. Dat wil zeggen, ze leren ons hele basale principes over de werking van ons brein. Mogelijk kunnen sommige van de bevindingen ook bijdragen aan het verbeteren van behandelingen voor bepaalde hersenziektes. Denk hierbij aan bipolaire stoornis (de kennis uit hoofdstuk 2), verslaving (hoofdstuk 6 en 7) of anorexia nervosa (hoofdstuk 9).

NOTE ON STATISTICS

Many of the experimental chapters contain references to a 'Supplementary statistics table'. These tables contain detailed test statistics (*t* and *F* values) as well as *P* values of all the comparisons that have been made in this thesis. In view of space, these supplementary statistics tables have been omitted from the printed version of this book, but will be made available together with publication of the chapters in a scientific journal.

LIST OF PUBLICATIONS

Published manuscripts

T.J.M. Roelofs, **J.P.H. Verharen,** G.A.F. van Tilborg, L. Boekhoudt, A. van der Toorn, J.W. de Jong, M.C.M. Luijendijk, W.M. Otto, R.A.H. Adan*, R.M. Dijkhuizen*. *A novel approach to map induced activation of neuronal networks using chemogenetics and functional neuroimaging in rats: A proof-of-concept study on the mesocorticolimbic system*. NeuroImage 156: 109-118 (2017)

J.P.H. Verharen, J.W. de Jong, T.J.M. Roelofs, C.F.M. Huffels, R. van Zessen, M.C.M. Luijendijk, R.Hamelink, I. Willuhn, H.E.M. den Ouden, G. van der Plasse, R.A.H. Adan*, L.J.M.J. Vanderschuren*. *A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states*. Nature Communications 9:731 (2018) OPEN ACCESS

J.P.H. Verharen, J. Kentrop, L.J.M.J. Vanderschuren*, R.A.H. Adan*. *Reinforcement learning across the rat estrous cycle.* Psychoneuroendocrinology 100: 27-31 (2019) OPEN ACCESS

** Shared senior authorship*

Manuscripts under revision

R. Wichmann, C.M. Vander Weele, A.S. Yosafat, E.H.S. Schut, **J.P.H. Verharen**, S. Sridharma, C.A. Siciliano, E.M. Izadmehr, K.M. Farris, C.P. Wildes, E.Y. Kimchi, K.M. Tye. *Acute stress induces long-lasting alterations in the dopaminergic system of female mice*. (Under revision, eLife) bioRxiv: https://doi.org/10.1101/168492

Chapters 1, 3, 4, 6, 7, 8 and 9 are being prepared for submission or are under review at a journal.


MIX
Paper from
responsible sources
FSC
www.fsc.org