TRANSLATIONAL PSYCHIATRY: WHEN AFFECT MOTIVATES EFFECT

DIRK GEURTS

ISBN/EAN:

978-94-6284-100-0

Cover design:

Mirthe Geurts

Text layout:

Dirk Geurts

Printed by:

Ipskamp Printing, Nijmegen

Translational Psychiatry: When affect motivates effect

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 4 mei 2017
om 14.30 uur precies

door

Dirk Everdina Maria Geurts

geboren op 9 december 1981
te Doetinchem

Promotoren

Prof. dr. R. Cools

Prof. dr. R.J. Verkes

Copromotor

Dr. H.E.M. den Ouden

Manuscriptcommissie

Prof. dr. B. Roozendaal

Prof. dr. O.A. van den Heuvel (VU Amsterdam)

Prof. dr. R.J. Heinz (Charité Universitätsmedizin Berlin, Duitsland)

CONTENTS

Part I

OPENING

1

GENERAL INTRODUCTION

In this introduction I first present the research aims and chapters in general terms. Second, I introduce Pavlovian to instrumental transfer (PIT), which is the central experimental psychological construct used in this thesis. Third, I provide a brief overview of the literature on the neural circuitry and chemistry of PIT, to serve as background for the empirical chapters. Fourth, an outline of this thesis is presented.

1.1 GENERAL INTRODUCTION TO THE AIMS OF THIS THESIS

Our actions are influenced by our affective states, or emotions (Damasio, 1994; Bechara et al., 2000; Dolan, 2002). Thus, a comprehensive understanding of why (and how) we do what we do, should involve a thorough understanding of how affective states modulate or motivate our actions. Here I focus on the motivational influence of affective states on instrumental, goal-oriented actions that lead us to get something (done).

Our understanding of the neurocognitive mechanisms underlying the influence of affective states on instrumental actions in humans is growing, but limited (Damasio, 1997; Cardinal et al., 2002; Brosch et al., 2013). Nevertheless, such understanding is particularly important because the majority of psychiatric disorders involve problems of affect (American psychiatric association, 2000). Putatively these disorders also involve aberrant affective influence on behaviour, be it too much or too little (Cools et al., 2008; Dayan and Huys, 2008; Dayan and Seymour, 2013; Heinz et al., 2016). Therefore, knowing how affect controls instrumental actions in patients suffering from these disorders may advance our understanding of psychiatric disorders and ultimately lead to more effective treatment regimes for those who suffer from them.

The general aim of this thesis is to advance our understanding of the neurocognitive and neurochemical mechanisms that underlie the motivational control of instrumental actions by affective cues in both healthy humans and those suffering from psychiatric disorders.

More specifically, the aim of this thesis is two-fold:

1. To increase our understanding of the neurocognitive and neurochemical mechanisms that allow affective states to (de)motivate instrumental actions in healthy humans; and

2. To leverage this knowledge to improve our understanding of the neurocognitive mechanisms underlying affective dysregulation of behaviour in patients with affective, impulsive and aggressive symptoms.

I address aim 1 in chapters 2 and 3. Specifically, I assessed the role of the neuromodulator serotonin on the motivational control of instrumental behaviour by affective cues in chapter 2. In chapter 3 I present a study in which I assessed activity and functional connectivity in the network of re-

gions subserving these motivational influences on behaviour, using functional magnetic resonance imaging (fMRI).

I address aim 2 in chapters 4-6, describing three cross-sectional fMRI studies in patients with borderline personality disorder (chapter 4) and psychopathic criminals (chapter 5 and 6) on the motivational control of instrumental behaviour by affective cues and associated neural underpinnings. In addition, the study presented in chapter 4 also included a one year follow-up assesment of borderline personality disorder symptom severity. This allowed me to assess whether symptom reduction in the patients with borderline personality disorder could be predicted based on the motivational control by affective cues.

1.2 THE INTERACTION BETWEEN AFFECT AND INSTRUMENTAL BEHAVIOUR

Adaptive behaviour depends on interactions between systems regulating affective versus rational control (Daw et al., 2005; Dayan et al., 2006; Evans, 2008; Huys et al., 2011). Such affective biases can help us respond adaptively to situations. For example, appetitive affect will motivate us to approach food or let us stay in the vicinity of nice people. In addition, negative affect might help us to be cautious (inhibit our behaviour for some time) or motivate us to withdraw from bad company. These biases can be found across phylogeny and are likely a product of an evolutionary selective process leading to mostly adaptive responses with regard to the environment in which they developed (Friston et al., 1994; Shettleworth, 2010).

These motivational influences of affect can also lead to irrational, but normal behaviour: Many decision-making phenomena that appear irrational, such as the framing effect (De Martino et al., 2006; Kahneman and Frederick, 2007) or the endowment effect (Kahneman et al., 1990; Morewedge and Giblin, 2015), may reflect the impact of affective cues on instrumental behaviour (Dayan et al., 2006; Dayan and Huys, 2008; Dayan and Seymour, 2013; Dayan, 2014). One of the most widespread 'artificial' motivational influences of affect on decision making is found in commercials: Advertisement companies are masters in coupling affective states to certain products, exploiting the motivational influence of affect on behaviour to lure one into buying. Think, for instance, of coupling an appetitive feeling (excited by happy families, sport and even Santa Claus) to a famous cola brand. This knowledge can, luckily, also be used for the good: Getting rid of the branded packages of cigarettes decreases tobacco-seeking (Hogarth et al., 2014). Furthermore, pictures that induce an aversive affect (e.g. of black

lungs) are used to inhibit us from buying or using cigarettes (Brewer et al., 2016).

In addition to understanding 'normal' behaviour, unraveling the motivational control of instrumental behaviour by affective cues is also pivotal for understanding several psychiatric symptoms and disorders: The majority of psychiatric patients have problems with affect and motivation (Salamone et al., 2015). Obviously these include patients suffering from the so-called affective disorders (i.e. mood and anxiety disorders). However problems with affective processing are not limited to these disorders. Many patients with psychotic disorders show changes in affect and motivation (consider for example negative symptoms in schizophrenia, Reddy et al., 2015). Moreover, aberrant affect is also central to personality disorders such as borderline personality disorder (American psychiatric association, 2000). Thus, affective problems play a major role in a vast variety of psychiatric disorders. However, how affective cues control instrumental behaviour in these disorders is relatively underexplored (see for review Brosch et al., 2013).

How aberrant affective processing leads to behavioural, psychiatric symptoms is mainly subject to hypothetical narratives about the role of affective processing in our lives, instead of rigorous empirical investigations. One important outstanding question is whether some of these disturbances of affect are related to the motivational influence of affect on on-going instrumental actions. This thesis is aimed to assess the latter hypothesis. Advancing our understanding of how phenomenological problems or symptoms of patients arise might not only provide us with new treatment perspectives, but might also help us leverage the efficiency of already existing treatment regimes. Such a translation from neurocognitive insights to (in the end) psychiatric patient care would be an instantiation of translational psychiatry. This thesis aimed to be part of such a translational approach.

The studies presented in this thesis represent well controlled, laboratory experiments aimed to advance our understanding of the mechanisms underlying the motivational control of instrumental behaviour by affective cues in healthy subjects and psychiatric patients. We have modeled the motivational control of instrumental behaviour in terms of an interaction between two distinct behavioural controllers: A Pavlovian controller to capture affective influences (Nees et al., 2015) and an instrumental controller to capture 'rational', goal-oriented behaviour (Dickinson and Balleine, 2002). Critically, in line with a long history of research, we refer to this interaction as Pavlovian to instrumental transfer (PIT) (see for review Holmes et al., 2010). In the next section we will describe these controllers and their interaction in more detail.

1.2.1 *Instrumental and Pavlovian conditioning*

Our actions in response to our environment are broadly speaking under control of two different systems: an instrumental and a Pavlovian system (Rescorla and Solomon, 1967; Dickinson and Balleine, 2002; Dolan and Dayan, 2013). Instrumental control refers to the fact that the organism can use actions to control outcomes in the environment (Dickinson and Balleine, 2002). The instrumental system builds upon learnt action-outcome (Adams and Dickinson, 1981; Balleine and Dickinson, 1998) and stimulus-action (Thorndike, 1911; Hull, 1943) associations. An organism typically learns instrumental associations when it executes a certain action (e.g. pressing) upon the appearance of a certain stimulus (e.g. a button) and finding that contiguous upon this action it attains a certain outcome (e.g. money) (see Figure 1.1 stage I). If this outcome or reinforcement is an appetitive one, the instrumental control system will be more likely to steer behaviour towards the same action when encountering the same stimulus again. Vice versa, if this reinforcement is an aversive one (e.g. monetary loss), the instrumental control system will be more likely to steer away from or inhibit the action. Adaptations of action that are based on the value of the outcome are observed when the instrumental behaviour is (mainly) under control of action-*outcome* associations, i.e. "goal-directed" control. Studies have shown that with extended instrumental training behaviour gradually tends to rely more on *stimulus*-action associations. Behavioural control based on these stimulus-action associations, i.e. behaviour under "habitual" control, is less sensitive to changes in the value of the outcome or changes in the contingency between action and outcome (Dolan and Dayan, 2013).

In contrast to instrumental control (thus relying on action-outcome and stimulus-action associations), Pavlovian control relies on stimulus-outcome learning, with Pavlovian responses representing evolutionarily preprogrammed, innate responses to an outcome-predictive stimulus (see Figure 1.1 stage II). Pavlovian conditioning involves the acquisition of affectively significant information: An originally neutral stimulus (the conditioned or conditional stimulus, CS) is presented together with an aversive or appetitive event (the unconditioned stimulus, US). The CS acquires aversive or appetitive properties and elicits a response (conditioned response, CR) that is often (but not always) similar to the response that individuals exhibit to the US (the unconditioned response UCR). For example, the animal may learn that the sound (CS) of food pellet delivery is associated with receiving these food pellets (US) irrespective of the behaviour that is displayed by the animal. The outcome or unconditioned stimulus (US) generates, depend-

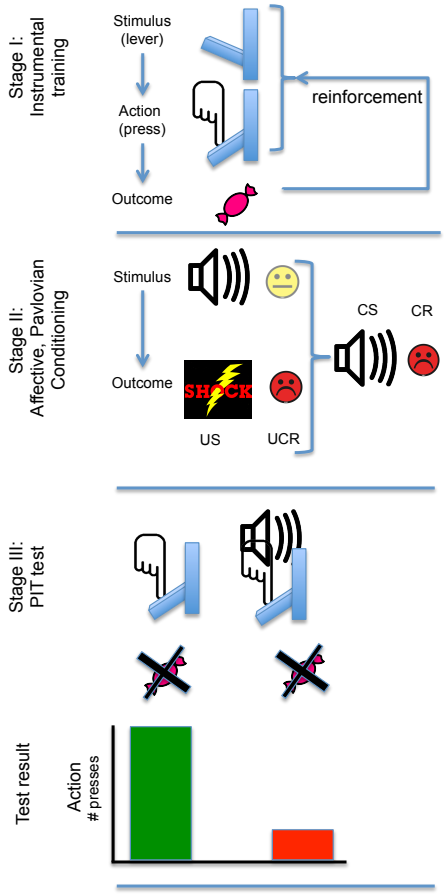


Figure 1.1 *– A typical aversive PIT paradigm. A typical aversive PIT paradigm consists of three stages. These stages ar depicted in the figure above including the expected PIT test results. Stage I: Instrumental learning leads to instrumental actions in reaction to the instrumental stimulus (lever). Stage II: Pavlovian conditioning with a shock (aversive unconditioned stimulus, US) leads to an aversive conditioned stimulus (CS, here a sound). Stage III: In the PIT stage the instrumental stimulus (the lever) is presented to the subject again, but now in extinction (i.e. without outcome delivery upon the appropriate action). Then during performance of the instrumental task, the CS is displayed. The PIT effect is the change in instrumental behaviour due to presentation of the CS. Test result: Here this is an attenuation of button presses due to display of an aversive (red) compared to a neutral (green) CS.*

ing on its nature, specific innate, untrained responses in the animal (e.g. approach, licking, saliva production, i.e. unconditioned responses (UCR)). These reactions, normally elicited by the US, are on the one hand coupled to the affective value of the US (i.e. preparatory reactions, e.g. vigorous approach and withdrawal to something good/desirable or bad/undesirable respectively) and on the other hand to specific characteristics of the US (i.e. consummatory reactions, e.g. watering mouth for specific food or eye-blink for an air puff in the eye). With respect to representing the affective value of the US by the CS, Pavlovian conditioning has been described as building an emotional or affective memory (Dolan, 2002). This makes studying effects of Pavlovian CS an interesting approach towards assessing affective changes. This approach has been exploited in the past decades (e.g. Estes and Skinner, 1941; Nees et al., 2015; Heinz et al., 2016). I will exploit specifically the affective reaction to the CS in this thesis.

Thus, the Pavlovian and instrumental system attribute value to aspects of our environment and actions respectively, which motivates our behaviour. More specifically, the instrumental system attributes value to certain actions in particular contexts, as these have been learnt to lead to certain, valued outcomes. These goal-directed actions can be regarded as a form of 'rational' behaviour in the sense that it is built on manipulating the environment in order to achieve a goal. The Pavlovian system endows previously neutral cues with affective value because of the predictive relationship with the affective properties of the predicted US. This system can be regarded as 'affective' in the sense that it amongst others elicits responses reflective of the affective value that has been coupled to the CS.

1.2.2 *Motivational control of instrumental actions by affective cues*

It is obviously advantageous to use information that is predictive of important future outcomes to guide ones actions. In general, adaptive responses can benefit from well-timed preparatory responses. Evolution endowed us with neural systems that underlie Pavlovian and instrumental learning and can take advantage of such predictive cues in our environment (Friston et al., 1994; Shettleworth, 2010). The Pavlovian and instrumental systems do not act in isolation. In most cases in our day-to-day live as well as in laboratory experiments their interaction is synergistic. For instance, consider the following laboratory procedure: A rat learns to press a lever in order to gain food in a standard instrumental paradigm. To be able to press the lever, the rat has to first approach the lever. During instrumental training, presentation of the lever (S) is not only associated with the action leading

to the outcome (S-A-O), but also directly with the outcome (S-O)(Yin et al., 2008). With this S-O association the lever will also become an appetitive Pavlovian CS, predicting an appetitive outcome. This appetitive CS elicits approach behaviour, thus facilitating the instrumental lever pressing.

However, it is hard to disentangle the contribution of the Pavlovian and instrumental system in situations where the synergy between these two systems prevails. Critically, there are also behaviours that are the result of Pavlovian-instrumental conflict (Dayan and Seymour, 2013). This antagonism shows on the one hand that indeed both systems are at work at the same time and on the other hand the relative strength of both systems. In the following I will describe some key experiments that evidence antagonism between these systems. I will end this section by describing the experimental design that we will use throughout this thesis: Pavlovian to instrumental transfer (PIT).

Most of the experiments that evidence antagonism between the instrumental and Pavlovian system reveal an interesting power of the affective Pavlovian system over the instrumental system: One excellent example comes from an experiment by Hershberger, who attempted to train chickens to move away from food in order to obtain it. In his experimental set-up a food bowl moved in the same direction as the chicken, but with twice the speed. Thus, when the chickens approached the food, the food bowl would rapidly move away from the chicken. The chickens were not able to learn (by action-outcome contingencies) to withdraw from the bowl in order to be able to eat the food. Their Pavlovian response to approach the bowl was too strong. This example illustrates an instance where the Pavlovian system maladaptively overrides the instrumental learning system in a seemingly hard-wired manner.

A recent series of experiments in humans uncovers another apparent conflict between the Pavlovian and instrumental system (for review see Guitart-Masip et al., 2014). These studies show that instrumental behaviour is biased by the Pavlovian system towards making active go-actions in pursuit of reward and towards making passive nogo responses to avoid punishment. People have difficulty being passive to obtain reward and being active to avoid punishment even when these actions would maximize positive outcomes. These natural couplings between motor activation and appetitive valence on the one hand and motor inhibition and aversive valence on the other hand thus bias the instrumental system (Dayan and Seymour, 2013). However, in the experimental set-up used in the above mentioned experiments Pavlovian and instrumental contingencies are not independent and

therefore the contribution of both systems cannot be readily segregated (see Swart et al., 2016).

A paradigm that makes use of the interaction between independently acquired Pavlovian cues and instrumental actions is Pavlovian to instrumental transfer (PIT, see Figure 1.1). This form of interaction between the Pavlovian and instrumental system exploits the motivational, affective importance of Pavlovian conditioned stimuli. PIT experiments typically consist of three stages: First, an instrumental training stage in which the subject learns an instrumental contingency (e.g. lever (S) → pressing the lever (A) → receiving money (O)). Second, a Pavlovian conditioning stage in which the subject learns a Pavlovian contingency (e.g. a sound (CS) → shock (US)). The third stage is effectively a combination of stage I and II: The subject is again shown the lever, which will elicit the instrumental actions learned in stage 1. This happens usually in extinction, i.e. without reinforcement of the instrumental response. Crucially, the Pavlovian conditioned stimulus (the sound) is presented concurrently with the instrumental stimulus. Thus, this stage allows for quantifying the motivational control of instrumental actions by affective cues. This can be measured as the change in instrumental behaviour induced by presentation of the affective Pavlovian CS (see Figure 1.1, stage III).

The PIT-procedure described in the previous paragraphs, is called 'outcome general' (as opposed to outcome-specific) PIT. The term outcome general refers to the critical property of this procedure, that affective properties of the CS (acquired through association with the US) lead to the PIT effect, but that this effect is not dependent on specific sensory properties of the US used during conditioning. Thus, two CS conditioned with different US with different sensory properties, but the same affective properties (e.g. juice and food for an equally hungry and thirsty participant) should elicit the same outcome-general PIT effect. The outcome general PIT paradigm is thus an operationalization of the general motivational control of instrumental behaviour by an affective cue. It provides an opportunity to measure and quantify the motivational influence of an affective cue with an empirical, experimental paradigm. This makes the general PIT paradigm especially suitable to achieve the aim of this thesis: To advance our understanding of the neurocognitive and chemical mechanisms that underlie the motivational control of instrumental actions by affective cues in both healthy humans and those suffering from psychiatric disorders. This motivational influence of affect on instrumental actions is thus specifically captured by the outcome-general PIT paradigm.

I note that there is another form of transfer that I did not study in this thesis: outcome-specific PIT. This form of transfer refers to the biasing of instrumental choice between two responses on the basis of sensory similarities between the outcome in instrumental training and the US in Pavlovian conditioning (Dickinson and Balleine, 2002). Thus, CS conditioned with a US that is similar to the outcome of a certain instrumental action, will bias choice towards this particular instrumental action. This kind of PIT has dissociable cognitive and neural underpinnings and is associated with more sensory specific (or consummatory) motivational effects on choice, rather than the general motivational (or preparatory) aspects of affective cues (Dickinson and Balleine, 2002). When I use PIT in this thesis we refer to outcome-general PIT unless otherwise specified.

One overarching aim of my thesis is to advance our understanding of the mechanisms underlying PIT. Most prior studies have been done in animals and were focused on appetitive PIT. In fact, only few human studies have assessed aversive PIT and mainly approach actions were considered in these studies. I will provide evidence in this thesis that focuses on aversive PIT in humans in approach as well as withdrawal behaviours.

Next I will review typical behavioural findings and current knowledge about the neural underpinnings of PIT (Holmes et al., 2010; Campese et al., 2015; Cartoni et al., 2016).

1.2.3 *Neural implementation of Pavlovian to instrumental transfer*

Behavioural findings

A broad literature has shown that CS associated with appetitive US invigorate instrumental behaviours (Cartoni et al., 2016), in contrast to inhibitory effects of CS paired with aversive US (Davis and Wright, 1979). However, these invigorating or inhibiting effects of CS do not appear to be universal, but critically depend on the type of instrumental behaviour: Aversive CS inhibit 'appetitive' behaviour (e.g. approach) (Estes and Skinner, 1941), but appear to invigorate aversive behaviour (e.g. avoidance or withdrawal) (Overmier et al., 1971; Campese et al., 2014). In contrast, appetitive CS invigorate appetitive behaviour (Estes, 1948; Corbit, 2005) and inhibit aversive behavior (Huys et al., 2011). The aforementioned findings have been replicated in humans (Di Giusto et al., 1974; Huys et al., 2011; Lewis et al., 2013; Lovibond and Colagiuri, 2013).

Recently a PIT paradigm was developed that encompasses all cells of Table 1.1 in a comprehensive paradigm (Huys et al., 2011). A study using this paradigm thus assessed the interaction between two different instru-

			Instrumental behaviour	
			Approach	**Withdrawal**
Pavlovian	**CS valence**	**Appetitive**	++ activation	-- inhibition
		Neutral	o no change	o no change
		Aversive	-- inhibition	++ activation

Table 1.1 *– Influence of different CS Valence on different forms of instrumental behaviour*

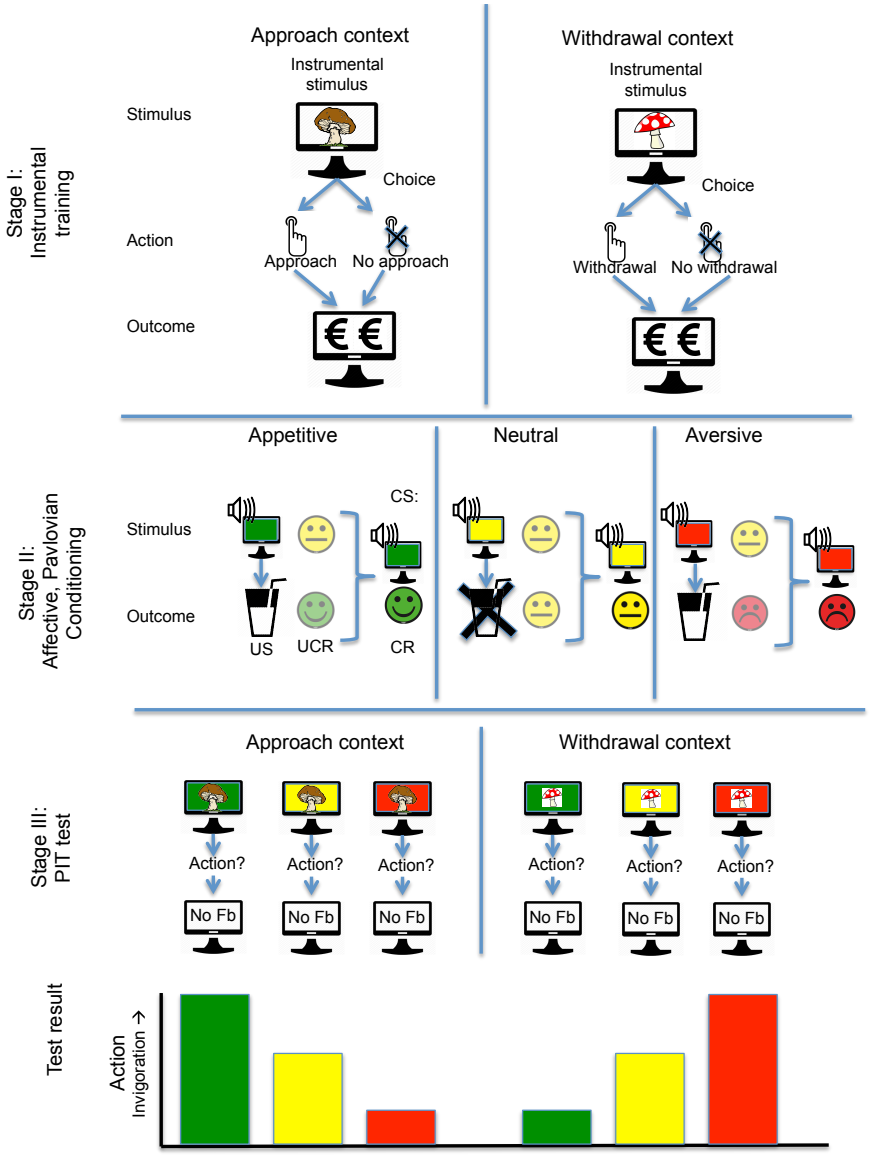
mental actions (approach/withdraw) with different Palvovian CS (appetitive/neutral/aversive). It revealed the expected effects (as depicted in Table 1.1, based on Rescorla and Solomon, 1967). In this thesis I also aimed to assess all the cells of Table 1.1 and transfer effects depicted in Box 1, with a specific focus on aversive CS valence.

The amygdala

The amygdala is the part of the brain most strongly implicated in emotional processing (Dolan, 2002; Phelps and LeDoux, 2005; Kim and Jung, 2006). In addition, the amygdala has recently been implicated in instrumental learning especially in the process of outcome valuation (Balleine and Doherty, 2009). The amygdala has been implicated in appetitive PIT in several animal (Hall et al., 2001; Holland and Gallagher, 2003; Corbit, 2005; Mahler and Berridge, 2011) and human studies(Talmi et al., 2008; Prevost et al., 2012): In animal studies lesions of the central nucleus of the amgydala abolished appetitive PIT (Corbit, 2005), whereas u-opioid stimulation of the amygdala enhanced appetitive PIT (Mahler and Berridge, 2011). Similarly, in humans, the first neuroimaging study that focused on PIT, also revealed a relation between amygdala BOLD signal and appetitive PIT on a subject-by-subject basis (Talmi et al., 2008). Another study showed that signal in the dorsal amygdala within the boundaries of the centromedial complex was involved in general appetitive PIT (Prevost et al., 2012). Moreover, a study by Campese et al (2014) showed that lateral as well as central amygdala and their functional connection were necessary for the expression of general aversive PIT in avoidance behavior in rats: Lesions of the lateral amygdala as well as lesions of the central amygdala (placed after instrumental and Pavlovian training) impaired the facilitation of instrumental avoidance behavior by aversive Pavlovian cues. Functional disconnection of the lateral and central amygdala also impaired this facilitation.

Box 1: Schematic depiction of the extended PIT task

The picture below depicts the typical three phases of a PIT task (cf. Figure 1.1). In addition, in this particular task we make use of two instrumental contexts and three different CS valence. This leads to a 2 x 3 design. The test effects expected based on the study of Huys et al are depicted at the bottom. Abbreviations: conditioned stimulus (CS), unconditioned stimulus (US), unconditioned respons (UCR), conditioned response (CR), No Feedback (No Fb).



Thus, there is evidence that implicates a key role for the central nucleus of the amygdala in appetitive and aversive general PIT. The main limitation of the human neuroimaging studies to date is that investigations were limited to appetitive instrumental behaviours. Thus a role for the amygdala in general PIT employing aversive instrumental behaviours (such as avoidance) still remains to be established.

Striatum: limbic-motor integration

Two key projection areas of the amygdala are the subcortical ventral striatum and dorsal striatum (including caudate and putamen). The ventral striatum has been described as serving a limbic-motor interface, through which motivation gets translated into - and putatively integrated with - action (Mogenson et al., 1980; Cardinal et al., 2002; Shiflett and Balleine, 2010). This region is indeed ideally positioned to integrate affective and instrumental information, because it receives abundant input from the amygdala as well as limbic regions of the prefrontal cortex and projects to structures involved in behavioural expression (Haber and Rauch, 2010). It is not surprising then that the nucleus accumbens is found to play a role in PIT. Indeed, the ventral striatum has been shown to be central to PIT: Ample evidence in rats show that general appetitive PIT in approach is disrupted by lesions of the nucleus accumbens core (Corbit et al., 2001; Borchgrave et al., 2002; Corbit and Balleine, 2011). Moreover, manipulation of dopaminergic transmission of the ventral striatum leads to changes in PIT (e.g. Dickinson et al., 2000; Corbit et al., 2007; El-Amamy and Holland, 2007). In humans Talmi et al.(2008) showed that the relation between nucleus accumbens signaling and instrumental behaviour was different during display of an appetitive CS compared to the display of a neutral CS, thereby suggesting a role for nucleus accumbens in appetitive PIT in humans.

Whereas the ventral striatum (esp. nucleus accumbens and ventral regions of caudate and putamen) is strongly implicated in Pavlovian processes (Salamone and Correa, 2012) and transfer processes, the dorsal striatum (caudate and putamen) is strongly implicated in instrumental behaviour (Balleine and Doherty, 2009). The caudate nucleus has abundant connections with the ventromedial prefrontal cortex and both structures are implicated in goal-directed control of behaviour (Yin et al., 2005a; 2005b). The posterior putamen, in contrast, is more involved in learning and maintaining habitual responses (Yin and Knowlton, 2006). A study by Corbit and colleagues (2007) suggested that integrity of the dorsolateral striatum in rats (analogue of the putamen in humans) was essential for PIT. There are no direct indications from animal work that the dorsomedial striatum is

involved in PIT. However, a human fMRI study employing a PIT task, that replaced the instrumental action with motor imagery, showed involvement of the dorsal and ventral parts of the caudate nucleus (Mendelsohn et al., 2014).

In summary, there is evidence that the amygdala, nucleus accumbens and the striatum are involved in general PIT. Nevertheless, only few (6) neuroimaging studies in human populations are available (Talmi et al., 2008; Prevost et al., 2012; Lewis et al., 2013; Ly et al., 2014; Mendelsohn et al., 2014; Garbusow et al., 2015) and, importantly, there is lack of paradigms investigating the role of especially aversive PIT for approach as well as withdrawal behaviours (cf. the different cells in Table 1.1; but see Lewis et al., 2013). In this thesis, I aimed to address this key lacuna.

1.2.4 *Neurochemical underpinnings of PIT: Role of serotonin*

The amygdala and nucleus accumbens receive abundant monoaminergic projections. One of the major monoaminergic neurotransmitters dopamine, a catecholamine related to reward processing and learning, has been found to play a critical role in appetitive PIT. Evidence from work with experimental animals indicates that general appetitive PIT is abolished by dopamine antagonists (Dickinson et al., 2000) and enhanced by dopamine agonists (Wyvell and Berridge, 2001). The ventral striatum appears key for these effects of dopamine. Infusing amphetamine (a dopamine agonist) into the nucleus accumbens enhanced appetitive PIT. Moreover, van Wassum et al. (2013) established a correlation between phasic dopamine release in the nucleus accumbens and behavioural general PIT. In addition, inactivation of the ventral tegmental area, a region with abundant dopamine projections to the accumbens, also attenuated general PIT (Murschall, 2006).

Another major monoaminergic neurotransmitter, serotonin, also plays a key role in PIT, but has received less attention. There has been a long debate about whether serotonin's primary function is better explained in terms of an affective, motivational account or in terms of behavioural inhibition (Soubrie, 1986; Deakin and Graeff, 1991). Recently, it has been proposed by several authors that serotonin's primary role might actually be best characterized as underlying the coupling between aversive affect and behavioural inhibition (Cools et al., 2008; Dayan and Huys, 2008; Crockett et al., 2009; Dayan and Huys, 2009; Boureau and Dayan, 2011; Cools et al., 2011). Such proposals were inspired by hypotheses that serotonin would serve as a motivational opponent to dopamine, which serves rather to couple reward with behavioural activation (Daw et al., 2002).

Aversive general PIT measures exactly this coupling between aversive affect and behavioural inhibition. Given the limited research on aversive PIT in general, it is not surprising that the available evidence on serotonin in averisve PIT in animals is very sparse, and exisiting results are mixed at best. Nonetheless, there are some suggestions that serotonin might indeed be involved in aversive PIT (Thiébot et al., 1980; Nielsen and Appel, 1985; but see Thiébot et al., 1984).

In chapter 2 of this thesis, I used an aversive PIT paradigm with the aim to settle this long standing debate about the role of serotonin in behavioural inhibition versus aversive processing (Faulkner and Deakin, 2014). If serotonin couples aversive valence to behavioural inhibition, thereby mediating aversive PIT, then attenuation of serotonergic processing would attenuate the inhibition of instrumental behaviour by aversive CS, i.e. aversive PIT. Alternatively, if serotonin reduces aversive processing per se, one might predict that reducing serotonergic transmission would enhance the impact of aversive PIT. Finally, if serotonin is primarily involved in behavioural inhibition, one would expect an increase in vigour irrespective of the presentation of a particular CS. I tested these hypotheses by assessing the influence of serotonergic transmission on PIT in chapter 2.

1.3 TRANSLATIONAL PSYCHIATRY: PATIENT STUDIES

The second general aim of this thesis was to use the experimental setup and findings of the first two chapters to explore whether such investigations could benefit our insight in patients with psychiatric disorders. Unfortunately our understanding of the mechanisms that lead to and underlie psychiatric disorders is poor. This lack of understanding reflects the heterogeneity of these disorders, the poor definition of their symptoms and underlying mechanisms. Like the major classification system in psychiatry (i.e. DSM, American psychiatric association, 2000), much research in psychiatry adopts relatively coarse behavioural measurement methods (e.g. questionnaires or clinical observations) that do not allow us to unravel the underlying mechanisms. Most classifications based on the DSM likely comprise multiple pathological processes. Indeed there is large individual variability, so that two patients classified with the same syndrome according to DSM, may exhibit hardly any overlap in terms of their behavioural deficits. In addition, effects of treatment can vary greatly across individuals with the same classification, suggesting different underlying mechanisms. Although the DSM approach has unmistakably improved communication between clinicians and unified research in psychiatry, it is time to move forward

to a more basic, mechanistic level of understanding. This is in line with the Research Domain Criteria (RDoC) initiative of the National Institute of Mental Health. According to the RDoC initiative (Cuthbert and Insel, 2013), common mechanisms that contribute to differing forms of psychopathology and their associated symptomatology may provide the basis for a new classification framework for research on mental disorders (Nees et al., 2015; e.g. Heinz et al., 2016).

As mentioned at the start of this introduction affective dysregulation is a core feature of many, if not all, psychiatric disorders. We argue that advancing the understanding and treatment of affective dysregulation in neuropsychiatry depends critically on progress in the cognitive neuroscience of affective value-based decision making. This is the case because affective responses involve the assessment of the affective value of environmental stimuli, which in turn have an important impact on ongoing decision making (Cardinal et al., 2002; Balleine and Doherty, 2009).

The increasing influence of cognitive neuropsychiatry is evidenced readily by ongoing debate about cognitive mechanistic approaches to psychiatric taxonomy (Casey et al., 2013; Cuthbert and Insel, 2013; Jones et al., 2015; Heinz et al., 2016). This debate is centered on the observation that most psychiatric disorders are spectrum disorders with each individual patient suffering from a unique constellation of symptoms.

In this thesis I argue that a translational cognitive neuropsychiatric approach (i.e. an attempt to bring insights from the cognitive neurosciences to psychiatric practice) is paramount for developing and optimizing treatment for these complex psychiatric disorders (cf. Jones et al., 2015; Heinz et al., 2016). Indeed, as is the case for most psychiatric disorders, symptoms of affective, impulsive and aggressive disorders have important cognitive-affective characteristics (Dayan and Seymour, 2013).

Here, I will focus on two patient groups: Patients with psychopathy and patients with borderline personality disorder. Aberrant impact of affect is apparent clinically in both of these impulsive and aggressive disorders (Linehan, 1993; Rosenthal et al., 2008; Blair, 2013). I adopted not only a categorical, group-wise approach, but also assessed, in line with dimensional psychiatry, individual differences within these patient groups: I asked whether there were any subject-by-subject associations between the motivational control of instrumental behaviour by affective cues and neural signatures on the one hand and specific clinical measures on the other hand.

1.3.1 *Borderline personality disorder*

Borderline personality disorder is characterized by affective dysregulation, impulsive behaviour and interpersonal hypersensitivity (Gunderson, 2011a). It is a debilitating psychiatric condition, affecting 1%-2% of the population. The functional impairment in patients is severe. This contributes to the high suicide rate of almost 10% (Black et al., 2004). Moreover, these patients constitute a disproportionately large subset of psychiatric in and outpatient populations, who consume considerably more (mental) healthcare resources than other psychiatric patients (Bender et al., 2001).

This personality disorder has long been thought to be chronic and untreatable. However, more recent data shows high remission rates (45% in 2 and 85% in10 years) with low relapse rates (15% in 10 years) (Gunderson, 2011b).

With respect to the treatment of borderline personality disorder, accumulating evidence highlights the clinical utility of several forms of psychotherapy. Dialectical behaviour therapy is the most frequently investigated comprehensive psychotherapy for borderline personality disorder (Linehan, 1993; Kliem et al., 2010). The primary indication of this specific form of psychotherapy is a DSM borderline personality disorder classification and it is targeted directly at resolving affective dysregulation. Multiple controlled randomized clinical trials, systematic reviews and a recent meta-analysis have shown clinically relevant changes in several domains of the borderline spectrum (e.g. suicides and suicide attempts, self-harming behaviour and general well-being). However, although these changes are found across the studied groups, treatment response is highly variable within the group and overall treatment effects are modest (Stoffers et al., 2012). Thus, 27-35% of patients continue to have admissions, self-harm and conduct suicidal gestures (Lana and Fernández-San Martín, 2013). This variability may arise from the heterogeneity of the patient population, and indeed from the 'disease' definition: According to the DSM (American psychiatric association, 2000) a patient has to fulfill at least 5 out of 9 phenomenological criteria to be classified as having borderline personality disorder. There are over 250 valid combinations and two BPD patients can have as few as a single common symptom. Additional sources of comorbidity including comorbid depression, anxiety of substance misuse disorders further increase the heterogeneity.

As argued above, neurocognitive mechanistic research might contribute to alleviating this problem of heterogeneity by characterizing core neurocognitive mechanisms that underlie the wide spectrum of phenomenological

characteristics. Such a mechanistic approach might also help us identify key predictors of treatment success and thus mitigate the large variability in treatment efficacy (Nitschke et al., 2009; Pizzagalli, 2010; Roiser et al., 2011; Garbusow et al., 2015; Jones et al., 2015; Månsson et al., 2015; Heinz et al., 2016; Huys et al., 2016). Critically, we currently have relatively little insight in the factors that contribute to the large individual variability in treatment success. With the study presented in chapter 5 we contribute to filling this gap.

We first investigated baseline neurobehavioral differences between healthy controls and patients with borderline personality disorder during the performance of an aversive PIT task. Previous research has shown enhanced processing of aversive affective stimuli in borderline personality disorder (e.g. Domes et al., 2006; Silbersweig et al., 2007). Accordingly, we hypothesized that patients with borderline personality disorder would exhibit excessive impact of aversive Pavlovian CSs on instrumental behaviour and associated neural responses relative to controls. Critically, we also assessed whether these responses predicted individual differences in symptom reduction 1 year after the start of treatment in the borderline personality disorder group.

1.3.2 *Psychopathy and criminality*

Roughly 1% of the general population has the phenomenological traits of a psychopath, but a far higher percentage of the jails are occupied by psychopathic criminals ($\pm$10-25 %) (Hare, 2003; Porter and Woodworth, 2006). This reflects the criminal tendencies associated with psychopathy (Porter and Woodworth, 2006; Blais et al., 2014). Unfortunately, there is currently no established treatment available.

Previous research on psychopathic criminals has particularly focused on affective processing per se, but it is unclear whether psychopathic criminals show an aberrant reaction to affective stimuli or not (Schultz et al., 2016). Here we go beyond this prior work by assessing how aversive affective cues alter instrumental behaviour and associated neural signals in psychopathic criminals, instead of looking at effects of aversive processing per se. Moreover, we also assess reward motivation and associated neural mechanisms.

Behaviour of psychopathic criminals is characterized by cold, instrumental behaviour especially targeted at egocentric goals (Hare, 2003). Their behaviour is hardly disturbed by discomfort of others, which is usually aversive to non-psychopathic people (Batson et al., 1987). Moreover, although ambiguous, laboratory studies suggest that psychopathic criminals might

be less sensitive to aversive affective CS (but see Arnett, 1997; Flor et al., 2002; Birbaumer et al., 2005; Schultz et al., 2016). As such, we hypothesized in contrast to patients with borderline personality disorder, that their behaviour might be characterized by too little behavioural inhibition in the face of aversive information. Moreover, psychopathy severity, in terms of the psychopathy checklist – revised (PCL-R) (Hare, 2003) scores, has been associated on a subject-by-subject basis with cerebro-spinal fluid measurements indicative of compromised serotonergic functioning (Soderstrom et al., 2001; 2003). In line with our hypothesis on serotonin, this led us to hypothesize that PCL-R would be related to aversive disinhibition on a subject-by-subject basis.

In addition, recent evidence suggests a central role for appetitive affective cues in impulsive/antisocial traits, which are central to the construct of psychopathy and relevant in predicting criminal behaviour (Buckholtz, 2010; Bjork et al., 2012). More specifically, functional MRI and PET evidence suggests that reward expectancy elicited by these cues might be key to understanding impulsive/antisocial traits (Buckholtz et al., 2010; Bjork et al., 2012). These seminal findings are however collected from healthy control, community samples and therefore preclude direct conclusions about its relevance for understanding overt criminality. In chapter 6 we fill this gap by translating the neurobiological underpinnings of reward expectancy in low and high impulsive/antisocial non-criminal groups to a psychopathic criminal group also scoring high on impulsive/antisocial traits. This furthers our understanding of the motivational effect of appetitive affect in relation to impulsive/antisocial traits and, critically, overt criminality on a neurobiological level. To assess reward expectancy in chapter 6 we departed from the PIT paradigm and used an adapted monetary incentive delay (MID) paradigm (Knutson et al., 2001; Rhein et al., 2015). I used this task because extending the findings based on the MID task of Buckholtz et al. (2010) to a clinical level would contribute to our aim to translate the motivational effect of affective cues to a clinical level. Indeed, the MID paradigm has been extensively used in healthy subjects and across diagnostic categories in psychiatric patient populations (Knutson and Heinz, 2015). A MID paradigm allows researchers to measure the neural impact of a cue that predicts a possible upcoming reward (see chapter 6 for a depiction of the MID task). Participants are asked to respond as quickly as possible to a target by pressing a button. Prior to this target, a cue indicates the possibility to gain a reward or not, after a button press within a given time window. This cue has been shown to reliably elicit appetitive affect and neural activity associated with this affect. This neural activity is found with such effect sizes that

it can be used in small samples (Knutson and Greer, 2008; Wu et al., 2014). Note that a difference between PIT and MID is that the appetitive cue in a MID task is not independently trained from the instrumental task and that the appetitive outcome of this cue is not independent of the actions of the participant. Thus the motivational effect of the affective cue in this task is unlikely purely Pavlovian.

Box 2: Functional magnetic resonance imaging
Functional magnetic resonance imaging (fMRI) enables us to indirectly measure neural processing in the brain. Indirectly, because with fMRI we do not measure activity of brain cells directly, but we measure a consequence of this activity: We measure changes in magnetic resonance signals that depend on the changes in hemoglobin bound oxygen concentration in the blood: The Blood-Oxygenation-Level-Dependent (BOLD) signal (Ogawa et al., 1990). Hemoglobin is a protein with magnetic properties that transports oxygen and is found in erythrocytes (i.e. red blood cells). The magnetic properties of hemoglobin differ depending on whether hemoglobin is oxygenated or not. The BOLD signal is stronger for oxygenated than deoxygenated hemoglobin. Activity of neurons is accompanied by the use of oxygen. In response to increased activity the blood flow in the vicinity of this activity increases and compensates for the oxygen consumption. This compensation is an overcompensation leading to a relative increase in the ratio oxygenated to deoxygenated hemoglobin. This results in an increased BOLD signal. Thus, BOLD signal is an indirect index of neural activity. Because the vascular response to increased neural activity occurs over the course of seconds, BOLD signal is considerably slower than the underlying neural responses. This gives fMRI a relatively low temporal resolution and one has to take this into account when developing and analyzing an fMRI/behavioural paradigm. fMRI has a relatively high spatial resolution (partly depending on the strength of the magnetic field) allowing a quite accurate localization (3-5 mm) of BOLD signal changes. Given that the brain is also active during rest, with fMRI we usually study activity in terms of a relative difference in BOLD between experimental conditions that only differ in the presence of the cognitive process under investigation. The functional connectivity method PPI (psychophysiological interaction) can be used to investigate changes in task-related correlation between BOLD signal in different brain regions. This analysis relies on the assumption that fMRI BOLD in one brain area can be explained by the interaction of a task related parameter with the BOLD response of another area (Friston et al., 1997). In chapter 3, 4, 5 and 6 we used such analyses to assess functional connectivity between frontal, cortical regions and subcortical limbic structures.

1.4 GENERAL APPROACH AND KEY QUESTIONS

To advance our understanding of the neurocognitive and chemical mechanisms that underlie the motivational influence of affect on instrumental actions I combined behavioural paradigms, psychopharmacology and neuroimaging in healthy controls and patient groups. I used two paradigms in this thesis: A PIT paradigm (chapter 2-5, Box 1) and a MID paradigm (MID, chapter 6). Furthermore, I used fMRI in chapter 3, 4, 5 and 6 to assess neural processing and connectivity during these tasks (see for a general description of fMRI Box 2). Moreover to attenuate serotonergic transmission in chapter 2 we used acute tryptophan depletion, which is described in Box 3. To increase our understanding of the neurocognitive and neurochemical mechanisms that allow affective states to (de)motivate instrumental actions in healthy humans we set-out to answer the following questions:

1. Does attenuation of serotonergic transmission lead to attenuation of especially the inhibiting effect of aversive affective cues on instrumental actions?

2. Are neural structures identified as contributing to (appetitive) PIT in humans and animals (i.e. amygdala and striatum) also involved in aversive PIT in humans?

3. With regard to the differential effects of affective cues on different instrumental behaviours (i.e. action specificity, see Table 1.1) we asked whether action specificity in Pavlovian control involves differential influences on neural regions that encode action specificity?

To improve our understanding of the neurocognitive mechanisms underlying affective dysregulation of behaviour in patients with affective, impulsive and aggressive symptoms I aimed to answer the following questions:

4. Do patients with borderline personality disorder show excessive aversive PIT on the behavioural and neural level compared with healthy controls?

5. Critically, from the perspective of translational psychiatry, can symptom reduction after one year of treatment be predicted based on aversive PIT-related BOLD-signal (especially in the amygdala)?

6. Do psychopathic criminals show reduced aversive PIT compared with healthy controls? Are individual differences in psychopathic severity accompanied by variation in aversive PIT and associated neural signaling?

7. Are enhanced neural responses to reward expectancy related to overt criminality in psychopathic criminals?

Box 3: Acute tryptophan depletion
Serotonin is a monoamine that cannot pass the blood-brain barrier and therefore the serotonin in our central nervous system has to be synthesized within the boundaries of this system. Serotonin in our nervous system is produced by troptophanhydroxylase 2 from its precursor tryptophan. Tryptophan is one of several large neutral amino acids (LNAA) that humans receive via dietary intake. Tryptophan (i) can pass the blood-brain barrier in competition with other LNAAs and (ii) just like other LNAAs tryptophan is used for protein synthesis in the liver. The acute tryptophan depletion procedure takes advantage of these two aspects of tryptophan to lower central serotonin availability: This procedure involves the oral administration of large quantities of LNAAs excluding tryptophan. Because protein synthesis in the liver will increase due to the higher concentration of LNAAs, more tryptophan will be used in this process and thus extract tryptophan from the circulation, causing a peripheral tryptophan depletion. In addition the competition over the blood-brain barrier further reduces the availability of tryptophan in the brain and thereby presumably leading to a decrease in serotonergic transmission.
Direct and indirect evidence suggests disturbed serotonergic transmission after ATD (Crockett et al., 2011; but see van Donkelaar et al., 2011). With regard to anatomical specificity of this effect it is important to realize that serotonin neurons (stemming from the raphe nuclei) innervate virtually all parts of the neocortex and substantial parts of subcortical structures. It might be that certain serotonergic projections (i.e. from the dorsal raphe nucleus) are more sensitive to ATD than other serotonergic projections (e.g. medial raphe nucleus projections), because these differ in physiologic properties (Faulkner and Deakin, 2014). A wealth of literature is available on the effects of ATD on several behavioural paradigms (see Faulkner and Deakin, 2014 for a focussed review).

1.5 OUTLINE OF THIS THESIS

Part I (chapter 2 and 3) of this thesis addresses the first aim of this thesis: To advance our understanding of the neural and neurochemical mechanisms that allow affective states to alter instrumental actions in healthy humans.

In chapter 2 I assessed how attenuation of serotonergic transmission affects the coupling of aversive and appetitive affective Pavlovian cues on

instrumental approach and withdrawal actions, using acute tryptophan depletion in healthy volunteers. Here I test the hypothesis that serotonin depletion attenuates the coupling between aversive affect and behavioural inhibition.

In chapter 3 I adjusted the task for use in the fMRI scanner. In this chapter I assessed the role of the amygdala, striatum and ventromedial prefrontal cortex in aversive PIT in approach and withdrawal. I hypothesized that the amygdala and ventral striatum would be involved in aversive PIT. Moreover, I expected that action specificity of the PIT effect would either be instantiated by direct Pavlovian modulation of regions that showed action specific BOLD-signal or by Pavlovian modulation of the connection between action-specific regions and regions that implement instrumental behaviour such as the striatum.

Part II (chapters 4, 5 and 6) of this thesis addresses the second aim of this thesis: To advance our understanding of the cognitive and neural mechanisms underlying affective dysregulation of behaviour in patients with affective, impulsive and aggressive symptoms.

In chapter 4 I investigated aversive PIT in patients with borderline personality disorder, using the same fMRI paradigm as in chapter 3. To establish the clinical relevance of our findings, I asked whether individual differences in PIT-related BOLD signal predict symptom reduction one year after the start of treatment: I hypothesized that, relative to healthy controls, BPD patients would exhibit excessive impact of aversive Pavlovian CSs on instrumental behaviour as well as altered BOLD signaling in the amygdala and fronto-striatal circuitry including the ventromedial prefrontal cortex and striatum. Moreover, I predicted that aversive PIT-related amygdala BOLD-signal in the BPD group would predict symptom reduction one year after the start treatment.

In chapter 5 I assessed aversive PIT in a sample of psychopathic criminals, again using fMRI. I hypothesized that aversive PIT would be reduced in psychopathy; second, that psychopathy would be accompanied by differential modulation of frontal and striatal brain regions, especially by regions processing aversive information such as the amygdala; and third, that the strength of this modulation would be accompanied by differences in aversive PIT on the behavioural level and psychopathic severity

In chapter 6, I additionally addressed appetitive motivation by employing a MID fMRI paradigm. I assessed whether psychopathic criminals show similar (or even greater) increases in ventral striatal reward expectancy-related BOLD signal as do (than) non-criminal healthy controls with high impulsive and antisocial traits (following Buckholtz et al., 2010). In addition,

overt criminality might only emerge in high-impulsive antisocial persons if the relatively high level of ventral striatal reactivity to reward expectation is not accompanied by appropriate regulation of other brain areas. I tested this latter hypothesis by assessing differences in neural, functional connectivity between the healthy control group scoring high on impulsive/antisocial traits and the criminal psychopathy group.

Finally, in chapter 7 the results of the previous chapters are summarized. I discuss my results, their limitations and their implication for future research.

1.6 REFERENCES

ADAMS CD, Dickinson A (1981) Instrumental responding following reinforcer devaluation. The Quarterly Journal of Experimental Psychology.

AMERICAN psychiatric association T (2000) Diagnostic and statistical manual of mental disorders, Fourth edition, Text Revision. Washington, DC: American Psychiatric Association.

ARNETT PA (1997) Autonomic responsivity in psychopaths: A critical review and theoretical proposal. Clinical Psychology Review 17:903–936.

BALLEINE BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407–419.

BALLEINE BW, Doherty JPOA (2009) Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. Neuropsychopharmacology 35:48–69.

BATSON CD, Fultz J, Schoenrade PA (1987) Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. J Pers.

BECHARA A, Damasio H, Damasio AR (2000) Emotion, decision making and the orbitofrontal cortex. Cereb Cortex 10:295–307.

BENDER DS, Dolan RT, Skodol AE, Sanislow CA, Dyck IR, McGlashan TH, Shea MT, Zanarini MC, Oldham JM, Gunderson JG (2001) Treatment utilization by patients with personality disorders. Am J Psychiatry 158:295–302.

BIRBAUMER N, Veit R, Lotze M, Erb M, Hermann C, Grodd W, Flor H (2005) Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. Arch Gen Psychiatry 62:799–805.

BJORK JM, Chen G, Hommer DW (2012) Psychopathic tendencies and mesolimbic recruitment by cues for instrumental and passively obtained rewards. Biological Psychology 89:408–415.

BLACK DW, Blum N, Pfohl B, Hale N (2004) Suicidal behavior in borderline personality disorder: prevalence, risk factors, prediction, and prevention. J Pers Disord 18:226–239.

BLAIS J, Solodukhin E, Forth AE (2014) A meta-analysis exploring the relationship between psychopathy and instrumental versus reactive violence. Criminal Justice and behavior.

BORCHGRAVE R, Rawlins JNP, Dickinson A, Balleine BW (2002) Effects of cytotoxic nucleus accumbens lesions on instrumental conditioning in rats. Exp Brain Res 144:50–68.

BOUREAU Y-L, Dayan P (2011) Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 36:74–97.

BREWER NT, Hall MG, Noar SM, Parada H, Stein-Seroussi A, Bach LE, Hanley S, Ribisl KM (2016) Effect of Pictorial Cigarette Pack Warnings on Changes in Smoking Behavior. JAMA Intern Med 176:905.

BROSCH T, Scherer K, Grandjean D, Sander D (2013) The impact of emotion on perception, attention, memory, and decision-making. Swiss Med Wkly.

BUCKHOLTZ JW, Treadway MT, Cowan RL, Woodward ND, Benning SD, Li R, Ansari MS, Baldwin RM, Schwartzman AN, Shelby ES, Smith CE, Cole D, Kessler RM, Zald DH (2010) Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. Nature Neuroscience 13:419–421.

CAMPESE VD, Kim J, Lázaro-Muñoz G (2014) Lesions of lateral or central amygdala abolish aversive Pavlovian-to-instrumental transfer in rats. Frontiers in Behavioral Neuroscience, 8, 161.

CAMPESE VD, Sears RM, Moscarello JM, Diaz-Mataix L, Cain CK, LeDoux JE (2015) The Neural Foundations of Reaction and Action in Aversive Motivation. 27:171–195.

CARDINAL RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience and Biobehavioral Reviews 26:321–352.

CARTONI E, Balleine B, Baldassarre G (2016) Accepted Manuscript. Neuroscience and Biobehavioral Reviews:1–62.

CASEY BJ, Craddock N, Cuthbert BN, Hyman SE, Lee FS, Ressler KJ (2013) DSM-5 and RDoC: progress in psychiatry research? Nat Rev Neurosci 14:810–814.

COOLS R, Nakamura K, Daw ND (2011) Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology 36:98–113.

COOLS R, Roberts AC, Robbins TW (2008) Serotoninergic regulation of emotional and behavioural control processes. Trends in Cognitive Sciences 12:31–40.

CORBIT LH (2005) Double Dissociation of Basolateral and Central Amygdala Lesions on the General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer. Journal of Neuroscience 25:962–970.

CORBIT LH, Balleine BW (2011) The General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer Are Differentially Mediated by the Nucleus Accumbens Core and Shell. Journal of Neuroscience 31:11786–11794.

CORBIT LH, Janak PH (2007) Inactivation of the Lateral But Not Medial Dorsal Striatum Eliminates the Excitatory Impact of Pavlovian Stimuli on Instrumental Responding. Journal of Neuroscience 27:13977–13981.

CORBIT LH, Janak PH, Balleine BW (2007) General and outcome-specific forms of Pavlovian-instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral tegmental area. European Journal of Neuroscience 26:3141–3149.

CORBIT LH, Muir JL, Balleine BW (2001) The role of the nucleus accumbens in instrumental conditioning: evidence of a functional dissociation between accumbens core and shell. The Journal of neuroscience, 21(9), 3251–3260.

CROCKETT MJ, Clark L, Robbins TW (2009) Reconciling the role of serotonin in behavioral inhibition and aversion: acute tryptophan depletion abolishes punishment-induced inhibition in humans. Journal of Neuroscience 29:11993–11999.

CROCKETT MJ, Clark L, Roiser JP, Robinson OJ, Cools R, Chase HW, Ouden HD, Apergis-Schoute A, Campbell-Meiklejohn D, Campbell-Meikeljohn D, Seymour B, Sahakian BJ, Rogers RD, Robbins TW (2011) Converging evidence for central 5-HT effects in acute tryptophan depletion. Mol Psychiatry 17:121–123.

CUTHBERT BN, Insel TR (2013) Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC medicine. http://doi.org/10.1186/1741-7015-11-126

DAMASIO A (1994) Descartes' Error. New York: Avon Books Trade Printing.

DAMASIO AR (1997) Neuropsychology. Towards a neuropathology of emotion and mood. Nature 386:769–770.

DAVIS H, Wright J (1979) Procedural and parametric variability in studies of conditioned suppression. Bulletin of the Psychonomic Society.

DAW ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. Neural Netw 15:603–616.

DAW ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature Neuroscience 8:1704–1711.

DAYAN P (2014) Rationalizable Irrationalities of Choice. Top Cogn Sci 6:204–228.

DAYAN P, Huys QJM (2008) Serotonin, Inhibition, and Negative Mood. PLoS Comput Biol 4:e4.

DAYAN P, Huys QJM (2009) Serotonin in affective control. Annu Rev Neurosci 32:95–126.

DAYAN P, Niv Y, Seymour B, D Daw N (2006) The misbehavior of value and the discipline of the will. Neural Networks 19:1153–1160.

DAYAN P, Seymour B (2013) Values and actions in aversion. In: Neuroeconomics: Decision making and the brain (Glimcher PW, ed). Academic Press.

DEMARTINO B, Kumaran D, Seymour B, Dolan RJ (2006) Frames, biases, and rational decision-making in the human brain. Science 313:684–687.

DEAKIN JF, Graeff FG (1991) 5-HT and mechanisms of defence. J Psychopharmacol (Oxford) 5:305–315.

DIGIUSTO JA, Di Giusto EL, King MG (1974) Heart rate and muscle tension correlates of conditioned suppression in humans. J Exp Psychol 103:515–521.

DICKINSON A, Balleine B (2002) The role of learning in the operation of motivational systems. In: Steven's handbook of experimental psychology: Learning, motivation and emotion, 3rd ed. (Gallistel CR, ed), pp 497–534. New York: John Wiley & Sons.

DICKINSON A, Smith J, Mirenowicz J (2000) Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. Behav Neurosci 114:468–483.

DOLAN RJ (2002) Emotion, Cognition, and Behavior. Science 298:1191–1194.

DOLAN RJ, Dayan P (2013) Goals and Habits in the Brain. Neuron 80:312–325.

DOMES G, Winter B, Schnell K, Vohs K, Fast K, Herpertz SC (2006) The influence of emotions on inhibitory functioning in borderline personality disorder. Psychol Med 36:1163–1172.

EL-AMAMY H, Holland PC (2007) Dissociable effects of disconnecting amygdala central nucleus from the ventral tegmental area or substantia nigra on learned orienting and incentive motivation. European Journal of Neuroscience 25:1557–1567.

ESTES WK (1948) Discriminative Conditioning .2. Effects of a Pavlovian Conditioned Stimulus Upon a Subsequently Established Operant Response. J Exp Psychol 38:173–177.

ESTES WK, Skinner BF (1941) Some quantitative properties of anxiety. J Exp Psychol 29:390.

EVANS JSBT (2008) Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. Annu Rev Psychol 59:255–278.

FAULKNER P, Deakin JFW (2014) Neuroscience and Biobehavioral Reviews. Neuroscience and Biobehavioral Reviews 46:365–378.

FLOR H, Birbaumer N, Hermann C, Ziegler S, Patrick CJ (2002) Aversive Pavlovian conditioning in psychopaths: peripheral and central correlates. Psychophysiol 39:505–518.

FRISTON KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6:218–229.

FRISTON KJ, Tononi G, Reeke GN, Sporns O (1994) Value-dependent selection in the brain: simulation in a synthetic neural model. Neuroscience, 59(2), 229–243.

GARBUSOW M et al. (2015) Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. Addiction Biology, 21(3), 719–731.

GEURTS DEM, Huys QJM, Ouden den HEM, Cools R (2013) Aversive Pavlovian Control of Instrumental Behavior in Humans. Journal of Cognitive Neuroscience 25:1428–1441.

GUITART-MASIP M, Duzel E, Dolan R, Dayan P (2014) Action versus valence in decision making. Trends in Cognitive Sciences 18:194–202.

GUNDERSON JG (2011a) Clinical practice. Borderline personality disorder. N Engl J Med 364:2037–2042.

GUNDERSON JG (2011b) Ten-Year Course of Borderline Personality Disorder. Arch Gen Psychiatry 68:827.

HABER SN, Rauch SL (2010) Neurocircuitry: a window into the networks underlying neuropsychiatric disease. Neuropsychopharmacology 35:1–3.

HALL J, Parkinson JA, Connor TM, Dickinson A, EVERITT BJ (2001) Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour. Eur J Neurosci 13:1984–1992.

HARE RD (2003) Manual for the Hare Psychopathy Checklist-Revised, 2nd ed. Toronto, Ontario, Canada: Multi Health Systems.

HEINZ A, Schlagenhauf F, Beck A, Wackerhagen C (2016) Dimensional psychiatry: mental disorders as dysfunctions of basic learning mechanisms. J Neural Transm, 123(8), 809–821.

HOGARTH L, Maynard OM, Munaf ò MR (2014) Plain cigarette packs do not exert Pavlovian to instrumental transfer of control over tobacco-seeking. Addiction 110:174–182.

HOLLAND PC, Gallagher M (2003) Double dissociation of the effects of lesions of basolateral and central amygdala on conditioned stimulus-potentiated feeding and Pavlovian-instrumental transfer. European Journal of Neuroscience 17:1680–1694.

HOLMES NM, Marchand AR, Coutureau E (2010) Pavlovian to instrumental transfer: A neurobehavioural perspective. Neuroscience and Biobehavioral Reviews 34:1277–1295.

HULL CL (1943) Principles of behavior. New York, USA: Appleton. Available at: http://psychclassics.yorku.ca/author.htm.

HUYS QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P (2011) Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding Rangel A, ed. PLoS Comput Biol 7:e1002028.

HUYS QJM, Gölzer M, Friedel E, Heinz A, Cools R, Dayan P, Dolan RJ (2016) The specificity of Pavlovian regulation is associated with recovery from depression. Psychol Med:1–9.

JONES KA, Menniti FS, Sivarao DV (2015) Translational psychiatry-light at the end of the tunnel. Annals of the New York Academy of Sciences 1344:1–11.

KAHNEMAN D, Frederick S (2007) Frames and brains: elicitation and control of response tendencies. Trends in Cognitive Sciences 11:45–46.

KAHNEMAN D, Knetsch JL, Thaler RH (1990) Experimental tests of the endowment effect and the Coase theorem. Journal of political Economy, 98, 1325–1348.

KIM JJ, Jung MW (2006) Neural circuits and mechanisms involved in Pavlovian fear conditioning: a critical review. Neuroscience and Biobehavioral Reviews, 30(2), 188–202.

KLIEM S, Kröger C, Kosfelder J (2010) Dialectical behavior therapy for borderline personality disorder: A meta-analysis using mixed-effects modeling. Journal of Consulting and Clinical Psychology 78:936–951.

KNUTSON B, Adams CM, Fong GW, Hommer D (2001) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. Journal of Neuroscience 21:RC159.

KNUTSON B, Greer SM (2008) Anticipatory affect: neural correlates and consequences for choice. Philosophical Transactions of the Royal Society B: Biological Sciences 363:3771–3786.

KNUTSON B, Heinz A (2015) Probing Psychiatric Symptoms with the Monetary Incentive Delay Task. BPS 77:418–420.

LANA F, Fernàndez-San Martín MI (2013) To what extent are specific psychotherapies for borderline personality disorders efficacious? A systematic review of published randomised controlled trials. Actas Esp Psiquiatr 41:242–252.

LEWIS AH, Niznikiewicz MA, Delamater AR, Delgado MR (2013) Avoidance-based human Pavlovian-to-instrumental transfer. Eur J Neurosci, 38(12), 3740–3748.

LINEHAN MM (1993) Cognitive-behavioral Treatment of Borderline Personality disorder. New York: NY Guilford Press.

LOVIBOND PF, Colagiuri B (2013) Facilitation of Voluntary Goal-Directed Action by Reward Cues. Psychological Science, 24(10), 2030–2037.

LY V, Cools R, Roelofs K (2014) Aversive disinhibition of behavior and striatal signaling in social avoidance. Social Cognitive and Affective Neuroscience 9:1530–1536.

MAHLER SV, Berridge KC (2011) What and when to "want?" Amygdala-based focusing of incentive salience upon sugar and sex. Psychopharmacologia 221:407–426.

MÅNSSON KNT, Frick A, Boraxbekk C-J, Marquand AF, Williams SCR, Carlbring P, Andersson G, Furmark T (2015) Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. Transl Psychiatry 5:e530.

MENDELSOHN A, Pine A, Schiller D (2014) Between Thoughts and Actions: Motivationally Salient Cues Invigorate Mental Action in the Human Brain. Neuron 81:207–217.

MOGENSON GJ, Jones DL, Yim CY (1980) From motivation to action: functional interface between the limbic system and the motor system. Progress in Neurobiology 14:69–97.

MOREWEDGE CK, Giblin CE (2015) Explanations of the endowment effect: an integrative review. Trends in Cognitive Sciences 19:339–348.

MURSCHALL A (2006) Inactivation of the ventral tegmental area abolished the general excitatory influence of Pavlovian cues on instrumental performance. Learning & Memory 13:123–126.

NEES F, Heinrich A, Flor H (2015) A mechanism-oriented approach to psychopathology: The role of Pavlovian conditioning. Int J Psychophysiol 98:351–364.

NIELSEN EB, Appel JB (1985) The effect of drugs on the acquisition of stimulus control in a conditioned suppression procedure. Psychopharmacology, 85(1), 80–86.

NITSCHKE JB Ph.D., Sarinopoulos I Ph.D., Oathes DJ Ph.D., Johnstone T Ph.D., Whalen PJ Ph.D., Davidson RJ Ph.D., Kalin NH M.D. (2009) Anticipatory Activation in the Amygdala and Anterior Cingulate in Generalized Anxiety Disorder and Prediction of Treatment Response. Am J Psychiatry 166:302–310.

OGAWA S, Lee TM, Nayak AS, Glynn P (1990) Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn Reson Med 14:68–78.

OVERMIER JB, Bull JA III, Pack K (1971) On insirumental response interaction as explaining the influences of Pavlovian CS+ s upon avoidance behavior. Learning and Motivation 2:103–112.

PHELPS EA, LeDoux JE (2005) Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. Neuron 48:175–187.

PIZZAGALLI DA (2010) Frontocingulate Dysfunction in Depression: Toward Biomarkers of Treatment Response. Neuropsychopharmacology 36:183–206.

PORTER S, Woodworth M (2006) Psychopathy and Aggression. In: Handbook of psychopathy, 1st ed. (Patrick CJ, ed), pp 481–494. New York: The Guilford Press.

PREVOST C, Liljeholm M, Tyszka JM, O'Doherty JP (2012) Neural Correlates of Specific and General Pavlovian-to-Instrumental Transfer within Human Amygdalar Subregions: A High-Resolution fMRI Study. Journal of Neuroscience 32:8383–8390.

BLAIR RJ (2013) Psychopathy: cognitive and neural dysfunction. Dialogues Clin Neurosci 15:181–190.

REDDY LF, Horan WP, Green MF (2015) Motivational Deficits and Negative Symptoms in Schizophrenia: Concepts and Assessments. In: Current Topics in Behavioral Neurosciences, pp 357–373 Current Topics in Behavioral Neurosciences. Cham: Springer International Publishing.

RESCORLA RA, Solomon RL (1967) Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychol Rev 74:151–182.

RHEIN von D, Cools R, Zwiers MP, van der Schaaf M, Franke B, Luman M, Oosterlaan J, Heslenfeld DJ, Hoekstra PJ, Hartman CA, Faraone SV, van Rooij D, van Dongen EV, Lojowska M, Mennes M, Buitelaar J (2015) Increased Neural Responses to Reward in Adolescents and Young Adults With Attention-Deficit/Hyperactivity Disorder and Their Unaffected Siblings. Journal of the American Academy of Child & Adolescent Psychiatry 54:394–402.

ROISER JP, Elliott R, Sahakian BJ (2011) Cognitive Mechanisms of Treatment in Depression. Neuropsychopharmacology 37:117–136.

ROSENTHAL MZ, Gratz KL, Kosson DS, Cheavens JS, Lejuez CW, Lynch TR (2008) Borderline personality disorder and emotional responding: A review of the research literature. Clinical Psychology Review 28:75–91.

SALAMONE JD, Correa M (2012) The mysterious motivational functions of mesolimbic dopamine. Neuron 76:470–485.

SALAMONE JD, Koychev I, Correa M, McGuire P (2015) Neurobiological basis of motivational deficits in psychopathology. European Neuropsychopharmacology 25:1225–1238.

SCHULTZ DH, Balderston NL, Baskin-Sommers AR, Larson CL, Helmstetter FJ (2016) Psychopaths Show Enhanced Amygdala Activation during Fear Conditioning. Front Psychol 7:903.

SHETTLEWORTH SJ (2010) Cognition, evolution, and behavior, 2nd ed. New York: Oxford University Press.

SHIFLETT MW, Balleine BW (2010) At the limbic-motor interface: disconnection of basolateral amygdala from nucleus accumbens core and shell reveals dissociable components of incentive motivation. European Journal of Neuroscience 32:1735–1743.

SILBERSWEIG D, Clarkin JF, Goldstein M, Kernberg OF, Tuescher O, Levy KN, Brendel G, Pan H, Beutel M, Pavony MT, Epstein J, Lenzenweger MF, Thomas KM, Posner MI, Stern E (2007) Failure of frontolimbic inhibitory function in the context of negative emotion in borderline personality disorder. Am J Psychiatry 164:1832–1841.

SODERSTROM H, Blennow K, Manhem A, Forsman A (2001) CSF studies in violent offenders: I. 5-HIAA as a negative and HVA as a positive predictor of psychopathy. J Neural Transm 108:869–878.

SODERSTROM H, Blennow K, Sjodin AK, Forsman A (2003) New evidence for an association between the CSF HVA: 5-HIAA ratio and psychopathic traits. Journal of Neurology, Neurosurgery, and Psychiatry 74:918–921.

SOUBRIE P (1986) Reconciling the role of central serotonin neurons in human and animal behavior. Behav Brain Sci 9:319–335.

STOFFERS JM, Völlm BA, Rücker G, Timmer A, Huband N, Lieb K (2012) Psychological therapies for people with borderline personality disorder. Cochrane Database Syst Rev 8:CD005652.

SWART JC, Froböse MI, Cook JL, Geurts DEM, Frank MJ, Cools R, Ouden den HEM (2016) Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. Submitted:1–34.

TALMI D, Seymour B, Dayan P, Dolan RJ (2008) Human Pavlovian Instrumental Transfer. Journal of Neuroscience 28:360–368.

THIÉBOT MH, Jobert A, Soubrie P (1980) Conditioned suppression of behavior: its reversal by intra raphe microinjection of chlordiazepoxide and GABA. Neurosci Lett 16:213–217.

THIÉBOT MH, Soubrie P, Hamon M, Simon P (1984) Evidence against the involvement of serotonergic neurons in the anti-punishment activity of diazepam in the rat. Psychopharmacology.

THORNDIKE EL (1911) Animal Intelligence. New York: Macmillan. Available at: http://psychclassics.yorku.ca/Thorndike/Animal/.

VAN Donkelaar EL, Blokland A, Ferrington L, Kelly PAT, Steinbusch HWM, Prickaerts J (2011) Mechanism of acute tryptophan depletion: is it only serotonin? Mol Psychiatry 16:695–713.

WASSUM KM, Ostlund SB, Loewinger GC, Maidment NT (2013) Phasic Mesolimbic Dopamine Release Tracks Reward Seeking During Expression of Pavlovian-to-Instrumental Transfer. Biological Psychiatry 73:747–755.

WU CC, Samanez-Larkin GR, Katovich K, Knutson B (2014) Affective traits link to reliable neural markers of incentive anticipation. NeuroImage 84:279–289.

WYVELL CL, Berridge KC (2001) Incentive sensitization by previous amphetamine exposure: increased cue-triggered "wanting" for sucrose reward. The Journal of neuroscience, 21(19), 7831–7840.

YIN HH, Knowlton BJ (2006) The role of the basal ganglia in habit formation. Nat Rev Neurosci 7:464–476.

YIN HH, Knowlton BJ, Balleine BW (2005a) Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. European Journal of Neuroscience 22:505–512.

YIN HH, Ostlund SB, Balleine BW (2008) Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. European Journal of Neuroscience 28:1437–1448.

YIN HH, Ostlund SB, Knowlton BJ, Balleine BW (2005b) The role of the dorsomedial striatum in instrumental conditioning. European Journal of Neuroscience 22:513–523.

Part II

NEUROCHEMICAL MECHANISMS UNDERLYING AFFECTIVE REGULATION OF BEHAVIOUR

2

SEROTONIN AND AVERSIVE PAVLOVIAN CONTROL OF INSTRUMENTAL BEHAVIOR IN HUMANS

Geurts, D. E. M., Huys, Q. J. M., Ouden, den, H. E. M., & Cools, R.

Journal of Neuroscience, 2013, 33(48), 18932–18939.

Adaptive decision-making involves interaction between systems regulating Pavlovian and instrumental control of behavior. Here we investigate in humans the role of serotonin in such Pavlovian-instrumental transfer in both the aversive and the appetitive domain using acute tryptophan depletion, known to lower central serotonin levels. Acute tryptophan depletion attenuated the inhibiting effect of aversive Pavlovian cues on instrumental behavior, while leaving unaltered the activating effect of appetitive Pavlovian cues. These data suggest that serotonin is selectively involved in Pavlovian inhibition due to aversive expectations and have implications for our understanding of the mechanisms underlying a range of affective, impulsive and aggressive neuropsychiatric disorders.

2.1 INTRODUCTION

Serotonin is implicated in healthy and disordered functions so wide ranging that elucidating its function is an important scientific puzzle. Best known are its contributions to aversive processing and behavioral inhibition, with evidence showing that a reduction in serotonin disinhibits behavior in the face of expected punishments (Tye et al., 1977; Soubrie, 1986; Graeff et al., 1996; Crockett et al., 2009; 2012). This work provided the basis for a recent proposal that serotonin has a specific role in tying aversive Pavlovian influences to instrumental inhibition (Dayan and Huys, 2008; Boureau and Dayan, 2011; Cools et al., 2011).

This proposal is grounded in a long history of psychological theory according to which there is a dichotomy of Pavlovian versus instrumental control of behavior. Instrumental behavior is elicited by learned associations of stimulus-action pairs with reinforcements, while Pavlovian behavior arises as reflexive responses to learned stimulus-associated outcome expectancies (Thorndike, 1911; Pavlov, 1927). Pavlovian and instrumental contingencies may act synergistically or competitively, and anomalous Pavlovian-instrumental interaction might be core to several neuropsychiatric disorders (Dayan et al., 2006; Dayan and Huys, 2008; Boureau and Dayan, 2011). For example, Dayan and Huys (2008; 2009) argue that serotonin deficiency, as seen in depression, leads to a failure to inhibit aversive thoughts and actions. Here we investigate whether and how serotonin regulates the coupling between aversive Pavlovian and instrumental control.

The paradigmatic example of such coupling is aversive Pavlovian-instrumental transfer (PIT), in which an aversive Pavlovian conditioned stimulus (CS) inhibits instrumental behavior (i.e. conditioned suppression; Huys et al., 2011; Geurts et al., in press). Effects of serotonin on aversive PIT have not been assessed in humans. We fill this gap, while also investigating the valence-specificity of serotonin's PIT effects. Specifically, we examined how aversive and appetitive PIT are affected by acute tryptophan depletion (ATD) -- a dietary procedure to deplete central serotonin levels in humans (Crockett et al., 2011). The hypothesis that the effect of ATD is particularly pronounced on aversive rather than appetitive PIT concurs with an accumulating body of theory and evidence (Soubrie, 1986; Deakin and Graeff, 1991; Daw et al., 2002; Dayan and Huys, 2008; Boureau and Dayan, 2011; Cools et al., 2011). However, there are also several studies suggesting a potential role in appetitive processing (Cools et al., 2005; Nakamura et al., 2008; Seymour et al., 2012). We aimed to resolve this discrepancy by conducting direct comparison between effects of ATD on aversive and appetitive PIT.

Finally we asked whether such effects are restricted to appetitive instrumental actions such as approach, or extend to aversive actions such as withdrawal. We have already shown that the effects of Pavlovian stimuli differ between instrumental approach and withdrawal (Huys et al., 2011). Serotonergic neurons densely innervate structures involved in active defensive behaviors (McNaughton and Corr, 2004; Saulin et al., 2011), raising the possibility that serotonin alters Pavlovian modulation of withdrawal as well as approach actions (Dayan and Huys, 2009; Boureau and Dayan, 2011).

2.2 MATERIALS AND METHODS

2.2.1 *Participants*

Fifty seven healthy right-handed volunteers (18-28 years old; mean age of 23.8 $\pm$ 2.8; 22 women) participated in this experiment. The study was approved by the local ethical committee at the Radboud University, Nijmegen. Participants were recruited via local advertisements, and screened during a screening session several days before the experiment for psychiatric and neurological disorders and MRI contra-indications by means of prescreening questionnaires and a (medical) interview by a trained physician. All volunteers gave written informed consent, and were paid for their participation. Exclusion criteria were any history of cardiac, hepatic, renal, pulmonary, neurological, psychiatric or gastrointestinal disorder, current medication use as well as first-degree family history of mood disorders.

We report data from 45 participants (18-28 years old; mean age of 23.8 $\pm$ 2.8), as 12 participants could not be included for the following reasons: Five participants did not tolerate the amino acid drink; one participant fainted during venepuncture; one participant did not return for the second session; data from two participants were lost due to technical errors; one participant reported not following the instructions. Two participants did not meet inclusion criteria for simple query trials during Pavlovian conditioning (see section 2.3.2).

2.2.2 *General Procedure*

Participants attended two test sessions at least 6 days apart (maximum 13), and were administered either a tryptophan depleting drink (TRP-) or a balanced amino acid drink (BAL) in a double-blind, placebo-controlled, crossover design (22 participants received TRP- and 23 received BAL on the first session). Prior to the test sessions, participants fasted overnight and low-

protein food was provided on the test days. Following a resting period of about 5.5 hours after drink intake (mean 5h24m, SD 12 min in the TRP- condition and mean 5h26m, SD 14 min in the balanced condition) to ensure stable and low TRP levels, participants performed a series of tasks. The task presented here was administered after another experiment involving fMRI scanning (to be reported elsewhere). The current experiment started approximately 7 hours after the amino acid drink intake (6h49m, SD 14 min in the TRP- condition and 6h55m, SD 20 min in the balanced condition).

Participants were seated comfortably in front of a personal computer with headphones. They used a mouse with their right hand to indicate their choices. Earnings were paid by bank transfer after the second session.

2.2.3 *Amino-Acid Mixtures*

Central tryptophan (TRP) was depleted by ingesting an amino-acid load that did not contain TRP but did include other large neutral amino acids (LNAAs) (Reilly et al., 1997). The quantities of amino acids in each drink were based on those used by Young et al (1985), though a 78.2 g mixture was employed to minimize nausea. Both amino-acid mixtures (prepared by Nutricia, Liverpool, UK) had the following ingredients: L-alanine, 4.1g; L-arginine, 3.7g; L-cystine, 2.0g; glycine, 2.4 g; L-histidine, 2.4 g; L-isoleucine, 6 g; L-leucine, 10.1 g; L-lysine, 6.7 g; L-methionine, 2.3 g; L-proline, 9.2 g; L-phenylalanine, 4.3 g; L-serine, 5.2 g; L-threonine, 4.9 g; L-tyrosine, 5.2 g; and L-valine, 6.7 g. The balanced amino drink contained additionally L-tryptophan, 3.0 g and the TRP- mixture 3.0 g of MCC filler. Female participants received a 20% reduction in quantity to account for lower average body weight. The drinks were prepared by stirring the mixture into approximately 200 ml tap water with a choice of lemon-lime or grapefruit flavouring to compensate for the unpleasant taste. Except for five of the excluded participants, participants reported no side effects apart from transient nausea following ingestion of the drink.

2.2.4 *Blood sample analyses*

Blood samples were taken twice, once before amino acid intake, and once prior to testing (9 min $\pm$5m), to establish the efficacy of the ATD procedure. Venous samples were taken in EDTA tubes, centrifuged at 2650 gmax during 20 min, and then pipetted into heparin aliquots. These were stored for a maximum of three weeks at -20°C before moving them to a -80°C environment. From there, they were sent to an external laboratory. Quantitative

amino acid analysis was performed by high-performance liquid chromatography as described elsewhere (Fekkes et al., 1995). The ratio total TRP / LNAA was calculated as 100 times the concentration of TRP divided by the summed concentrations of the LNAAs (TRP/ΣLNAA) (Fernstrom and Wurtman, 1972).

2.2.5 *Task description*

We used the task as previously described in (Huys et al., 2011). In short, the task was divided into two blocks (approach and withdrawal), each consisting of an instrumental training, a Pavlovian training and a PIT stage (Figure 2.1).

Instrumental training

The instrumental task (Figure 2.1A) was an approach or withdrawal go/nogo task, framed in terms of collecting or discarding mushrooms. In the approach block, participants chose whether to collect the mushroom by moving the mouse towards and clicking on the stimulus (approach-go) within a response-window of 1.5 seconds, or not collect the mushroom by abstaining from a response for 1.5 seconds (approach-nogo). In the withdrawal block, participants chose whether to discard mushrooms by clicking in a blue frame located on the opposite side of the stimulus (withdrawal-go) or do nothing (withdrawal-nogo). The outcome (+/- 20 Euro cents) was then presented in the middle of the screen. Reinforcements were probabilistic, with the 'correct' response for each mushroom leading to gain or avoidance of loss on 75% of the trials. Correct trials were those on which participants discarded a 'bad' or kept a 'good' mushroom, and those on which they collected a 'good' or refrained from collecting a 'bad' mushroom. Participants had to learn the better response for each stimulus from the noisy reinforcement feedback. There were 3 'good' and 3 'bad' mushrooms in each context, meaning that all actions (i.e. approach-go, approach-nogo, withdrawal-go and withdrawal-nogo) could be followed by both rewards and punishments (Figure 2.1 and Table 2.1). Thus, the expected value of correct approach and withdrawal actions were equal and positive on average.

Pavlovian training

The second part of the task consisted of a separate classical conditioning procedure. Five compound Pavlovian stimuli (CS), consisting of a fractal visual stimulus (Figure 2.1B) and a tone were deterministically paired with

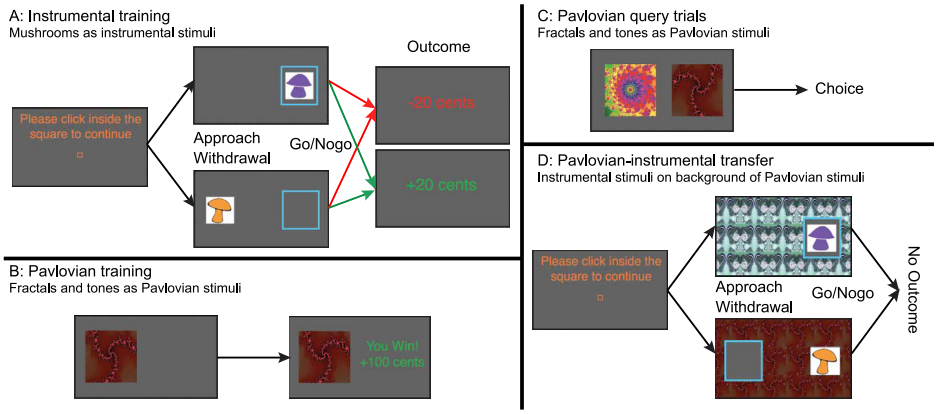


Figure 2.1 – A. Instrumental training. *To centre the cursor, participants clicked in a central square. The experiment consisted of a block with exclusively instrumental approach trials (n=120) and a block with exclusively withdrawal trials (n=120). In approach trials (top), participants chose whether to move the cursor towards the mushroom and click inside the blue frame onto the mushroom (go), or do nothing (nogo). In withdrawal trials, they instead moved the cursor away from the mushroom and clicked in the empty blue frame (go) or did nothing (nogo). Outcomes were presented immediately after go actions, or after 1.5 seconds. Per block there were 3 "good" and 3 "bad" instrumental stimuli. Participants played each block ones per testing day. Instrumental stimuli were different for both blocks, but the same for both days.* ***B. Pavlovian conditioning.*** *Participants passively viewed stimuli and heard auditory tones, followed by wins and losses. There were five fractal/tone combinations. Each combination was displayed 12 times in the first block and another 6 times in the second block.* ***C. Pavlovian query trials*** *On Pavlovian query trials, participants chose between two Pavlovian stimuli. No outcomes were presented, but they were counted and added to the total presented at the end of the experiment. Query trials were administered after every five Pavlovian conditioning trials.* ***D. Pavlovian-instrumental transfer.*** *Participants responded to the instrumental stimuli trained during the instrumental training stage, with Pavlovian stimuli tiling the background. No outcomes were presented, but participants were instructed that their choices counted towards the final total. No explicit instructions about the contribution of Pavlovian stimuli towards the final total were given.*

Block	Type of Instrumental Stimulus ('mushroom')	If the following action:	Then the following outcome (75%/25%):
Approach	Go ('good')	Go ('collect')	+20/-20
		Nogo ('avoid')	-20/+20
	Nogo ('bad')	Go ('collect')	-20/+20
		Nogo ('avoid')	+20/-20
Wihtdrawal	Go ('bad')	Go ('throw away')	+20/-20
		Nogo ('collect')	-20/+20
	Nogo ('good')	Go ('throw away')	-20/+20
		Nogo ('collect')	+20/-20

Table 2.1 *– Action outcome contingencies for the different instrumental stimuli*

outcomes. The best (S^P_{++}) and worst (S^P_{--}) CSs predicted a gain / loss of 100 cents while the intermediate CSs (S^P_{+},S^P_{0},S^P_{-}) were followed, respectively, by [+10 0 -10] cents. To ensure that participants paid attention, a query trial was presented on every fifth trial. Participants then had to choose between two different Pavlovian stimuli (Figure 2.1C) in extinction.

Pavlovian-Instrumental transfer

This was the main part of interest. Subjects chose whether to collect or discard mushrooms while the Pavlovian stimuli tiled the entire background (Figure 2.1D). Critically, no outcomes were presented. Participants were instructed to continue performing the instrumental task; that choices were still earning them the same outcomes and were being counted; but that they would not be told about the outcomes.

2.2.6 *Psychometric measurements*

Participants completed the following questionnaires during the screening session: Barratt Impulsiveness Scale (Patton et al., 1995), Behavioral Inhibition/Behavioral Activation Scale (Carver and White, 1994), Beck Depression Inventory (Beck et al., 1996), Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975), Hamilton Rating Scale for Depression (Hamilton, 1960), Spielberger Trait Anxiety Inventory (Spielberger et al., 1983), Kirby Questionnaire (Kirby and Maraković, 1996), Sensitivity to Punishment and Sensitivity to Reward Questionnaire (Torrubia et al., 2001) and the Dutch reading test (Schmand et al., 1991). Scores are reported in Table 2.2. In addition, participants completed the Positive Affect Negative Affect Schedule (PANAS) (Watson et al., 1988) and the Bond and Lader Visual Analogue Scales (BLV) (Bond and Lader, 1974) just before the PIT experiment. Finally, a neuropsy-

chological test battery was administered at the end of each testing day (approximately 15 minutes after the end of the PIT experiment) consisting of a number cancellation task, a box completion test, and a digit span (Table 2.2).

2.2.7 *Data analysis*

Data were analysed using the statistic software SPSS 16.0. Where appropriate, we performed repeated-measures analyses of variance (rmANOVA). Factors included Drink (2 levels: BAL/TRP-, within-subject), Order (2 levels: TRP- first or BAL first, between-subjects) and other factors defined below.

Serum levels

The TRP/ΣLNAA ratio was used as the dependent variable in an rmANOVA with Time (2 levels: Before/After, within-subject) x Drink x Order. This was followed by simple effects analyses: an rmANOVA with factor Time for each drink separately, and an rmANOVA with factor Drink only to comparing TRP/ΣLNAA after ATD and BAL.

Pavlovian conditioning

The threshold for performing above chance at the query trials (Figure 2.1C) was set to at least 14 (out of 18) correct (based on a sign test). Proportion correct choices were also submitted to a Drink x Order rmANOVA.

Instrumental training

There were four trial types, consisting of stimulus for which the correct response was: (i) go-approach, (ii) nogo-approach, (iii) go-withdrawal, (iv) nogo-withdrawal. We calculated the proportion of correct responses (p(correct)) for the first and last 10 trials of each trial type, both for the instrumental training and for the PIT stage. To assess whether participants learned to make the correct choice during instrumental training, we used an rmANOVA with Time (2 levels: first/last trial bin), Action Context (2 levels: approach/withdrawal), Correct Choice (2 levels: go/nogo), Drink and Order.

To assess whether the learned behavior generalized to the PIT stage, the two level factor Time was changed to include 3 levels (henceforth "extended Time factor"): the last instrumental, the first PIT and the last PIT trial bin.

Questionnaire	BAL1st	TRP-1st
Barratt -Total	59.4 (3.1)	54.3 (4.1)
Barratt-Attention	16.1 (1.0)	14.4 (1.1)
Barratt-Motor	18.9 (1.1)	17.6 (1.5)
Barratt-Non Planning	24.4 (1.3)	22.2 (1.7)
BIS	18.3 (0.8)	17.0 (0.9)
BAS-Total	24.7 (1.1)	28.1 (3.1)
BAS-Reward	8.9 (0.5)	9.1 (0.5)
BAS-Drive	7.6 (0.5)	7.5 (0.5)
BAS-Fun	8.2 (0.3)	8.4 (0.5)
BDI	1.1 (0.35)	1.4 (0.36)
EPQ-Psychoticism	2.1 (0.4)	1.2 (0.2)
EPQ-Extraversion	10.0 (0.4)	9.3 (0.6)
EPQ-Neuroticism	2.2 (0.4)	1.9 (0.3)
EPQ-Lie	6.2 (0.6)	6.6 (0.7)
HRSD	0.9 (0.2)	0.9 (0.3)
STAI	30.2 (1.3)	30.5 (1.2)
Kirby-Small	0.04 (0.01)	0.05 (0.02)
Kirby-Medium	0.03 (0.01)	0.03 (0.02)
Kirby-Large	0.02 (0.01)	0.03 (0.02)
SPSRQ-Punishment	4.9 (0.7)	4.4 (0.6)
SPSRQ-Reward	11.7 (0.8)	10.8 (0.9)
NLV	85.8 (1.4)	85.6 (1.8)
Number cancelation	227.7 (6.2)	207.5 (5.7)
Box completion	79.7 (3.0)	73.5 (3.6)
Digit span	16.2 (0.6)	18.1 (0.6)

Table 2.2 *– Trait characteristics and data from neuropsychological background tests as a function of Order (BAL1st/TRP-1st) (standard errors of the mean). Abbreviations: Barratt, Barratt Impulsivity Scale; BAS, behavioral activation system score; BDI, Beck Depression Inventory; BIS, behavioral inhibition system score from the BIS/BAS scale; EPQ, Eysenck Personality Questionnaire; Kirby, Kirby Questionnaire; NLV, Dutch reading test; SPSRQ, Sensitivity to Punishment and Sensitivity to Reward Questionnaire; STAI, Spielberger Trait Anxiety Inventory; HRSD, Hamilton Rating Scale for Depression*

Adequate generalization to the PIT stage implies an absence of any effect of, or interaction with, the factor Time.

Pavlovian-instrumental transfer stage

The primary effect of interest was that of the Pavlovian CSs on instrumental responding. This was assessed in terms of choice (the proportion of 'go' responses (p(go)) as a function of CS Valence and Action Context. We analysed this using an rmANOVA with Drink, Action Context and CS Valence (5 levels: S^P_{++} /S^P_{+} /S^P_{0} /S^P_{-} /S^P_{--}, within-subject). We modeled CS Valence as a linear contrast (+2/+1/0/-1/-2).

Planned contrasts were targeted at the most aversive and most appetitive Pavlovian stimuli (i.e. S^P_{++} and S^P_{--}). For this analysis, the five-level factor Pavlovian Valence in the omnibus rmANOVA was replaced by a Pavlovian Valence factor with 2 levels: i.e. S^P_{++} and S^P_{--}.

To account for variability of no interest introduced by the cross-sectional design (two days) and the blocked design of the PIT task (two blocks) we added the following between-subject factors to the rmANOVAs described above. First, to capture variance due to test-retest effects from day 1 to day 2 we added a between-subject factor Order (started with BAL on day 1 [BAL1st]/started with TRP- on day 1 [TRP-1st]). Note that the interaction between Order and the within-subject factor Drink (BAL/TRP-) is statistically similar to a main effect of Day (day1/day2). Likewise, in order to capture variability of no interest in PIT task performance that might be caused by Block Order (i.e. better performance on the second compared to the first block) we added a between-subject factor First Block (approach as first block [appr1st]/withdrawal as first block [wthd1st]). An interaction between Block Order and the within-subject factor Action Context (Approach/Withdrawal) represents a main effect of Block (block1/block2).

2.3 RESULTS

2.3.1 *Blood plasma analysis*

Acute tryptophan depletion (ATD) resulted in decreased TRP/ΣLNAA ratio as evidenced by a significant Drink x Time interaction ($F_{(1,41)}$ =492.9, p<.001) (Table 2.3). This was due to a 92.8% decrease in the TRP/ΣLNAA ratio following TRP-. The TRP/ΣLNAA ratio was lower for the TRP- than the BAL condition at the start of the PIT experiment ($F_{(1,41)}$=866.4, p<.001).

	TRP-		BAL	
	Before	**After**	**Before**	**After**
Day 1	9.44 (0.33)	0.63 (0.08)	9.04 (0.27)	5.80 (0.18)
Day 2	9.53 (0.31)	0.73 (0.08)	9.42 (0.28)	6.83 (0.37)

Table 2.3 *– Values present TRP/ΣLNAA ratio before and after drink ingestion on day 1 and day 2 (standard errors of the mean).*

2.3.2 *Pavlovian conditioning*

Participants performed highly accurately on the query trials during the Pavlovian stage evidencing successful Pavlovian conditioning (mean p(correct) = .97 correct, SD=.04). Two participants performed at chance level and were removed from further analysis. There was no significant effect of ATD on accuracy on the query trials (mean p(correct)BAL=.97, mean p(correct)TRP-=.97, $F_{(1,41)}$ =0.54, p=.82).

2.3.3 *Instrumental responding*

Participants showed robust acquisition of the instrumental contingencies (Figure 2.2; main effect of Time $F_{(1,41)}$ =242.4, p<.001) and this effect was maintained throughout the PIT stage (no main effect of, or interaction with, the extended Time factor; all ($F_{(1,40)}$ /(2,80)<3.0, p>.091). ATD impaired instrumental learning (Drink x Time interaction $F_{(1,41)}$ =4.9, p=.033; mean p(correct) at the end of the instrumental training: TRP-: .77, BAL: .82; mean improvement in p(correct) between first and last stage of instrumental learning: TRP-: .22; BAL: .28). This effect was maintained in the PIT stage (main effect of Drink: $F_{(1,40)}$ =6.7, p=.014; no interaction with extended Time factor).

ATD also specifically impaired nogo-approach actions (Figure 2.2): there was a significant three-way Drink x Action Context x Correct Choice interaction ($F_{(1,41)}$ =5.7, p=.022) which was driven by a main effect of Drink on nogo-approach stimuli ($F_{(1,41)}$ =10.7, p=.002) with the effect of ATD on all other actions failing to reach significance ($F_{(1,41)}$<1.2). Thus, ATD impaired the ability to make nogo responses in order to passively avoid bad mushrooms.

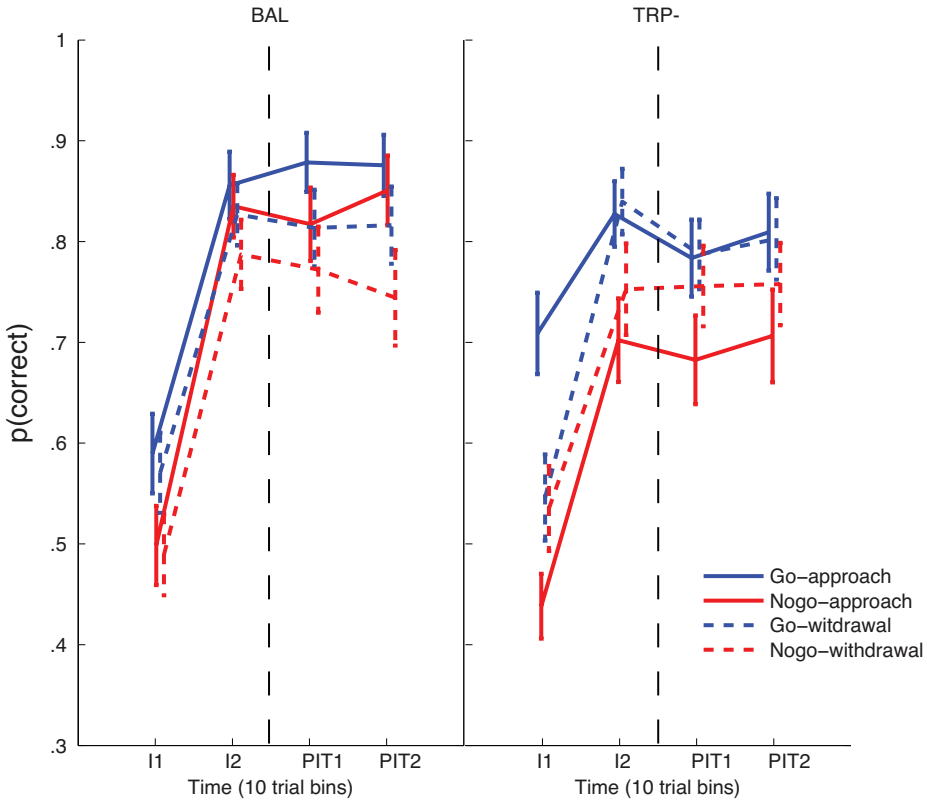


Figure 2.2 *– Instrumental learning and generalization to the Pavlovian-instrumental transfer stage after tryptophan depletion (right graph, TRP-) and after the balanced amino acid drink (left graph, BAL). The proportion of correct choices (p(correct)) are divided over the four different types of instrumental stimuli: go-approach, nogo-approach, go-withdrawal and nogo-withdrawal. Time is represented by bins of 10 trials for each type of stimulus at the beginning (I1) and the end (I2) of the instrumental training and at the beginning (PIT1) and the end (PIT2) of the PIT stage Error bars represent standard errors of the mean.*

2.3.4 *Pavlovian-instrumental transfer*

Consistent with our primary hypothesis, ATD altered the effects of Pavlovian stimuli on instrumental behavior. Specifically, the inhibitory effect of aversive Pavlovian CSs on instrumental responding seen at baseline was reversed by ATD (Drink x CS Valence (5 levels: S^P_{++} /S^P_{+} /S^P_{o} /S^P_{-} /S^P_{--}) linear contrast $F_{(1,41)}$=4.3, p=.045; planned contrast Drink x CS Valence (2 levels: S^P_{++} /S^P_{--}) $F_{(1,41)}$ =5.5, P=.023). This effect was driven by an effect of ATD on the aversive CSs. For the aversive CSs, the proportion of go choices was larger after TRP- than after BAL (main effect of drink for S^P_{--} only: $F_{(1,41)}$ =6.8, p=.013). Responding to the appetitive CSs was unaltered by ATD ($F_{(1,41)}$ =0.1, p=.74). Thus, ATD abolished the inhibitory effect of aversive Pavlovian CSs on instrumental responding.

The order of Drink sessions was counterbalanced, so that 22 participants received TRP- on the first session, and 23 received TRP- on the second session. Moreover, the analyses reported above were conducted with testing order as a between-subjects factor to account for variability of no interest. Contrary to our expectation, this factor interacted with the effect of interest. There was a significant three-way interaction between Drink x CS Valence (S^P_{++} /S^P_{--}) x Order ($F_{(1,41)}$ =11.7, p=.001, Figure 2.3). Breakdown of this three-way interaction effect by group (BAL1st versus TRP-1st) revealed that our effect of interest, i.e. aversive disinhibition after ATD, was present in participants who received BAL, but not in those who received TRP- on day 1 (Drink x CS valence for BAL1st: $F_{(1,21)}$=18.8, p<.001; for TRP-1st: $F_{(1,20)}$=0.5, p=.49, Figure 2.3). Furthermore, alternative break down of the interaction effect by Day (day 1 versus day 2) revealed that it was present on the first, but not the second day (Day 1: Drink x CS valence: $F_{(1,41)}$ =5.9, p=.02, Day 2: $F_{(1,41)}$ <0.1, p=.88).

As in the overall group, these effects in the BAL 1st group were also driven by the aversive CSs rather than the appetitive CSs. Thus, in this BAL 1st group, for the aversive CSs, the proportion of go choices was larger after TRP- than after BAL (main effect of drink for S^P_{--} only: $F_{(1,21)}$ =13.9, p=.001), while responding to the appetitive CSs was unaltered by ATD ($F_{(1,21)}$ =1.7, p=.21). Moreover, there was also a significant interaction between Drink, CS Valence in this BAL 1st group when comparing the aversive with the neutral CSs, ($F_{(1,21)}$=6.7, p=.017), but not when comparing the appetitive with the neutral CSs ($F_{(1,21)}$=2.1, p=.16). This interaction is depicted in Figure 2.3 and confirms a specific effect of ATD on aversive PIT.

In supplementary analyses we assessed whether this aversive disinhibition in the BAL 1st group was due to increased proportion of go choices

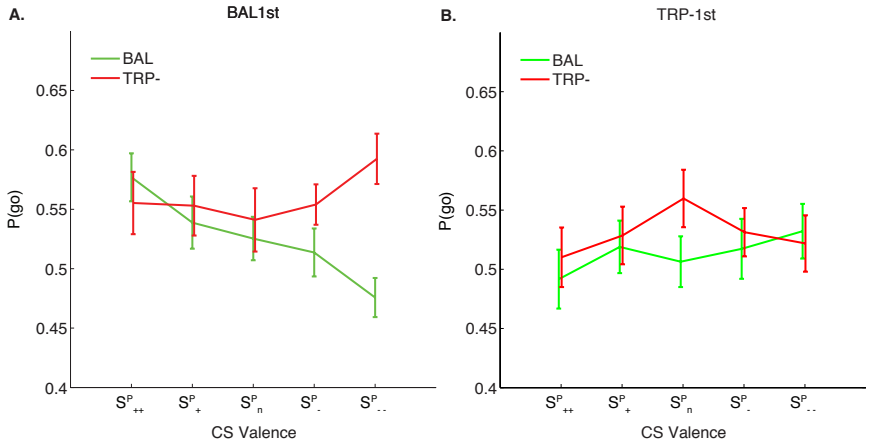


Figure 2.3 *– Behavioral data form the PIT stage as function of group. Shown are choice data as a function of CS Valence (S^P_{++} /S^P_{+} /S^P_{0} /S^P_{-} /S^P_{--}) after acute tryptophan depletion (TRP-, red line) and after the balanced amino acid drink (BAL, green line).* ***A****, Participants who started with BAL on day 1.* ***B****, Participants who started with TRP- on day 1. Error bars represent SEM.*

when go was correct, when go was an error, or some combination of both. To this end we assessed the probability of correct responses in the PIT stage during the aversive trials only using an ANOVA with an additional within-subject factor Type of Stimulus. There were two types of stimuli, one that required a go and one that required a nogo response to be correct. This analysis revealed a significant Drink x Type of Stimulus interaction for the aversive CSs ($F_{(1,20)}$=8.4, p=.009), which was due to increased proportion of correct Go responses after TRP- versus BAL ($F_{(1,21)}$=9.2, p=.007). There was no effect of ATD on the proportion of correct Nogo responses ($F_{(1,21)}$=0.9, p=.773). Thus, the aversive disinhibition induced by serotonin depletion was driven by increased proportion of go choices when go was correct and not when go was an error.

With respect to the proportion of go choices, we did not find a significant interaction between Action Context (approach versus withdrawal) and CS Valence (cf. Huys et al., 2011) across sessions ($F_{(1,41)}$ =.6, p=.43) or after BAL only ($F_{(1,41)}$ <.1, p=.95) and no modulation of this interaction by Drink ($F_{(1,41)}$ =1.3, p=.27) (Table 2.4).

	BAL1st				TRP-1st			
	BAL		TRP-		BAL		TRP-	
	Appr	Wthd	Appr	Wthd	Appr	Wthd	Appr	Wthd
S^P_{++}	.581 (.162)	.572 (.163)	.590 (.124)	.521 (.157)	.453 (.151)	.525 (.141)	.538 (.175)	.477 (.175)
S^P_{+}	.550 (.156)	.528 (.157)	.576 (.116)	.530 (.168)	.515 (.139)	.518 (.141)	.537 (.170)	.516 (.160)
S^P_{0}	.560 (.126)	.491 (.204)	.609 (.141)	.473 (.139)	.435 (.130)	.573 (.116)	.553 (.172)	.561 (.165)
S^P_{-}	.543 (.172)	.485 (.161)	.559 (.151)	.549 (.170)	.482 (.164)	.548 (.108)	.513 (.115)	.546 (.151)
S^P_{--}	.492 (.164)	.459 (.179)	.606 (.159)	.579 (.146)	.490 (.152)	.570 (.119)	.511 (.159)	.528 (.194)

Table 2.4 *– Values represent proportion of go actions as a function of Order (BAL1st/TRP-1st), Drink (BAL/TRP-), Action Context (approach/withdrawal) and CS Valence (very appetitive (S^P_{++}) to very aversive (S^P_{--})) during the Pavlovian-instrumental transfer stage (standard errors of the mean).*

2.3.5 *Order effects*

The order effect might raise the concern that random assignment of participants to groups failed, resulting in differences between groups (BAL1st versus TRP-1st) in vulnerability to ATD. Therefore we investigated whether there was evidence for any differences between the groups with respect to screening questionnaires and background neuropsychological tests (Table 2.2). The only measure that differed between the groups and was not affected by Drink or Day was the digit span test: Participants who received BAL on day 1 performed more poorly on the digit span task across both sessions than participants who received TRP- on day 1 (main effect of Order: $F_{(1,41)}$ =6.4, p=.015). However, adding this measure as a covariate in the omnibus rmANOVA did not reduce significance of the interaction between Order, Drink, and CS Valence (Order x Drink x CS Valence (S^P_{++}/S^P_{--}): $F_{(1,40)}$=7.7, p=.008), and it also did not interact with our main finding of aversive disinhibition (Digit Span x Drink x CS Valence: $F_{(1,40)}$=0.4, p=.511).

2.3.6 *Mood ratings*

Positive affect as measured with the PANAS immediately prior to the PIT experiment was significantly affected by ATD ($F_{(1,37)}$=9.7, p=.004, Table 2.5). Critically, this effect was not related to our main finding, i.e. no correlation existed between the effects of ATD on positive affect and the effects of ATD on the inhibiting effect of the aversive Pavlovian cue (Pearson $r_{(41)}$=-0.11, p=.51). In addition, we did not find any other main effect of or interaction with ATD on the other mood ratings (BLV subscales: $F_{(1,43)}$<1, p>0.52, PANAS negative affect: $F_{(1,37)}$=0.3,p=.58). Thus the finding that ATD modu-

	BAL	TRP-
PANAS-positive	25.4 (0.8)	27.7 (1.1)
PANAS-negative	12.4 (0.9)	11.8 (0.4)
BLV-alertness	45.1 (1.0)	44.8 (1.0)
BLV-contentedness	48.1 (1.8)	47.0 (1.2)
BLV-calmness	55.2 (0.7)	54.5 (1.0)

Table 2.5 *– Mood ratings as a function of Drink (BAL/TRP-) (standard errors of the mean). Abbreviations: BLV: Bond and Lader visual analogue scale; PANAS: Positive Affect Negative Affect Schedule.*

lates the inhibitory impact of an aversive Pavlovian stimulus is unlikely to be mediated by ATD-related changes in mood.

2.4 DISCUSSION

Results show that serotonin depletion attenuates aversive Pavlovian-instrumental transfer (PIT) without affecting appetitive PIT, thus providing evidence for a selective role of serotonin in tying aversive expectations to behavioral inhibition. This concurs with current theories according to which serotonin serves as a motivational opponent to dopamine (Daw et al., 2002; Boureau and Dayan, 2011; Cools et al., 2011). According to these theories, both serotonin and dopamine have coordinated effects that serve to couple a motivational axis (appetitive versus aversive processing), and an activational axis (energizing versus inhibiting behavior). In contrast to dopamine, which is well established to promote behavioral activation to seek rewards, serotonin was hypothesized to inhibit actions when punishment may occur. Data from the PIT phase of the current study concur with this hypothesis. The supplementary finding that withholding an action in the approach context is compromised by ATD also fits with this framework. Moreover it generally concurs with rodent work showing that performance on passive avoidance tasks is particularly vulnerable to manipulations that lower serotonin transmission, while leaving active avoidance unaltered (Soubrie, 1986).

To appreciate the relevance of PIT it is important to recognize that instrumental learning always involves both instrumental as well as Pavlovian contingencies (Yin et al., 2008). Therefore, PIT might influence the majority of instrumental responses, and these influences might be core to a wide range

of adaptive and maladaptive behaviors (Dayan et al., 2006; Guitart-Masip et al., 2012; Huys et al., 2012). Consider the specific cases of depression and impulse control disorders. Both implicate low serotonin, an observation that appears paradoxical given that depression has been primarily associated with aversive abnormalities, while impulsivity has been associated primarily with behavioral disinhibition (Cools et al., 2008a). The present data strengthen the hypothesis that serotonin does not control aversive processing per se or behavioral inhibition per se, but rather facilitates the coupling between aversive processing and behavioral inhibition. Accordingly, serotonin deficiency, as seen in depression and impulsivity, is accompanied not by enhanced impact of aversive stimuli per se or by reduced inhibition per se, but rather by reduced impact of aversive stimuli on the inhibition of behavior (as well as thoughts)(Dayan and Huys, 2008; Crockett et al., 2009; Boureau and Dayan, 2011; Cools et al., 2011; Huys et al., 2012; Robinson et al., 2012).

The first empirical evidence in humans for this hypothesis came from work by Crockett et al. (2012), who used a reinforced categorization task rather than a PIT task to show that ATD abolishes slowing of responding in the presence of punishment-predicting stimuli. The present study extends this work, not least by enabling direct comparison of aversive with appetitive Pavlovian influences. We show that the effects of ATD are valence-specific, and are restricted to the aversive domain. This observation concurs with some classic accounts of serotonin, according to which it is involved in aversive rather than appetitive processing (Deakin, 1983; Deakin and Graeff, 1991; but see Kranz et al., 2010). Moreover, it fits with formal theories, according to which serotonin is involved in the aversive side of model-free learning (Daw et al., 2002). Third, it is consistent with our previous findings, showing that ATD altered performance on a punishment, but not reward prediction learning task (Cools et al., 2008b; Robinson et al., 2012). Specifically, we have shown that ATD enhanced the ability to predict punishment while leaving reward prediction unaffected (Cools et al., 2008b). Initially, we interpreted this effect to reflect enhanced punishment prediction learning (Cools et al., 2008b). However, the present finding suggests that these prior observations might reflect disinhibition of responding in anticipation of punishment rather than enhanced punishment learning (cf. Dayan and Huys, 2009; Robinson et al., 2012).

The observation that our effects were restricted to the aversive domain might not seem consistent with electrophysiological data, revealing reward-responsive neurons in the dorsal raphe nucleus, the primary source of serotonergic input into the brain (Nakamura et al., 2008; Bromberg-Martin et

al., 2010; Okada et al., 2011). However, it should be recognized that the dorsal raphé nucleus contains a number of different types of nonserotonergic units that are likely to also be recorded. Thus the serotonergic identity of these neurons is not known. In addition, there are also serotonin depletion studies in marmosets and humans emphasizing effects in the reward domain (Rogers et al., 2003; Cools et al., 2005; Man et al., 2011) (maybe reflecting interactions with dopamine). For example, we have shown that ATD abolished speeding of responding with increasing feedback likelihood in a monetary incentive delay like task (Cools et al., 2005). The findings of the latter study may be reconciled with the present observation by recognizing that the ATD-induced abolition of speeding in that study might have resulted not just from reduced sensitivity to reward, but also from enhanced sensitivity to punishment. Of course, we acknowledge that our findings do not exclude effects of serotonin on reward processing outside the domain of PIT, for example in the domain of delayed discounting (Miyazaki, 2012).

The effects on aversive PIT are unlikely to reflect attenuation by ATD of Pavlovian conditioning per se or instrumental conditioning per se. First, we did not observe any effects on the query trials during the Pavlovian stage, although we acknowledge that this might not be the most sensitive measure. Second, the pattern of performance on the PIT stage is not consistent with an attenuation of Pavlovian conditioning. Attenuation of Pavlovian conditioning would have led to a flattening rather than a reversal of PIT effects. Third, effects on the PIT stage are also not confounded by effects during the instrumental learning stage. We did find effects of ATD during instrumental learning, with general declines of learning as well as a specific passive avoidance deficit after the depleting drink. However, these effects cannot account for the PIT effect, because the latter was restricted to the aversive domain, not extending to the appetitive domain.

One caveat of the present study is that the effect of interest was present only in the group of participants that received the balanced drink (BAL) on day 1 (Figure 2.3) (although it was significant when both groups were collapsed). Those who received the tryptophan depleting drink on day 1 did not show an effect of ATD. In fact these participants, who also exhibited greater working memory capacity, did not show any PIT effect at all, even when tested after BAL. Thus an unexpected result was the absence of PIT after BAL in half of our participants, who incidentally also had greater working memory capacity (as measured with the digit span). We consider two possibilities. First participants with greater working memory capacity might be less vulnerable to Pavlovian response biases. This account is less plausible given the lack of a continuous association between working mem-

ory capacity and our effect of interest. Alternatively, the combined administration of the ATD and an affective manipulation that induces a certain cognitive/affective state (e.g. a PIT task) might lead to transfer or reinstatement of that state from a first testing session to subsequent testing sessions, even though the subsequent testing sessions were not done under ATD. According to this account the abolition of PIT after BAL on day 2 may reflect formation of an association between abolished PIT and reduced serotonin states during the first visit. This alternative account concurs generally with the associative hypothesis of recurrence in depression (Robinson and Sahakian, 2008) and with empirical data from Robinson et al. (2009). They showed that negative mood induction under ATD led to negative mood after ATD on a second day. Critically on this second day there was no mood induction and the effect was not found for participants who received BAL on the first day.

A final point is that we had expected, based on Huys et al. (2011) that the aversive Pavlovian stimuli would influence instrumental responses in an action-specific manner, inhibiting approach-go actions and promoting withdrawal-go actions. We did not replicate these effects and consider the following accounts: The key difference with the paradigm of Huys et al. (2011) is that we explicitly modulated monoamines in our participants and that food intake was restricted during several hours before the experiment. This resulted in a drop in the TRP/ΣLNAA ratio even after the balanced amino acid drink (30%). It might well be that this relatively small drop in TRP/ΣLNAA might have been sufficient to disrupt action-specificity of PIT. This speculation concurs with the abolition of action-specificity in pathologies associated with serotonergic dysfunction, such as depression (Q.J.M. Huys et al, unpublished data).

In conclusion, these data suggest that serotonin is selectively involved in Pavlovian inhibition due to aversive expectations. These findings might have implications for our understanding of the mechanisms underlying a range of affective, impulsive and aggressive neuropsychiatric disorders, which have been associated with abnormal serotonin transmission. An obvious next step would be to assess the putatively aberrant Pavlovian biases on instrumental behavior in these patient groups, to advance our understanding of the neurochemical and cognitive mechanisms underlying these disorders.

2.5 REFERENCES

BECK AT, Steer RA, Ball R, Ranieri W (1996) Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. J Pers Assess 67:588–597.

BOND A, Lader M (1974) The use of analogue scales in rating subjective feelings. British Journal of Medical Psychology 47:211–218.

BOUREAU Y-L, Dayan P (2011) Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 36:74–97.

BROMBERG-MARTIN ES, Hikosaka O, Nakamura K (2010) Coding of task reward value in the dorsal raphe nucleus. Journal of Neuroscience 30:6262–6272.

CARVER CS, White TL (1994) Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. J Pers Soc Psychol 67:319.

COOLS R, Blackwell A, Clark L, Menzies L, Cox S, Robbins TW (2005) Tryptophan Depletion Disrupts the Motivational Guidance of Goal-Directed Behavior as a Function of Trait Impulsivity. Neuropsychopharmacology 30:1362-1373.

COOLS R, Nakamura K, Daw ND (2011) Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology 36:98–113.

COOLS R, Roberts AC, Robbins TW (2008a) Serotoninergic regulation of emotional and behavioral control processes. Trends in Cognitive Sciences 12:31–40.

COOLS R, Robinson OJ, Sahakian B (2008b) Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. Nature 33:2291–2299.

CROCKETT MJ, Clark L, Apergis-Schoute AM, Morein-Zamir S, Robbins TW (2012) Serotonin Modulates the Effects of Pavlovian Aversive Predictions on Response Vigor. Neuropsychopharmacology 37:2244-2252.

CROCKETT MJ, Clark L, Robbins TW (2009) Reconciling the Role of Serotonin in Behavioral Inhibition and Aversion: Acute Tryptophan Depletion Abolishes Punishment-Induced Inhibition in Humans. Journal of Neuroscience 29:11993–11999.

CROCKETT MJ, Clark L, Roiser JP, Robinson OJ, Cools R, Chase HW, Ouden HD, Apergis-Schoute A, Campbell-Meiklejohn D, Campbell-Meikeljohn D, Seymour B, Sahakian BJ, Rogers RD, Robbins TW (2011) Converging evidence for central 5-HT effects in acute tryptophan depletion. Mol Psychiatry 17:121–123

DAW ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. Neural Netw 15:603–616.

DAYAN P, Huys QJM (2008) Serotonin, Inhibition, and Negative Mood. PLoS Comput Biol 4:e4.

DAYAN P, Huys QJM (2009) Serotonin in affective control. Annu Rev Neurosci 32:95–126.

DAYAN P, Niv Y, Seymour B, D Daw N (2006) The misbehavior of value and the discipline of the will. Neural Networks 19:1153–1160.

DEAKIN JF (1983) Roles of serotonergic systems in escape, avoidance and other behaviors. In: Theory in Psychopharmacology (S C, ed). London and New York: Academic Press.

DEAKIN JF, Graeff FG (1991) 5-HT and mechanisms of defence. J Psychopharmacol (Oxford) 5:305–315.

EYSENCK HJ, Eysenck SBG (1975) Manual of the Eysenck Personality Questionnaire. London: Hodder and Stoughton.

FEKKES D, van Dalen A, Edelman M, Voskuilen A (1995) Validation of the determination of amino acids in plasma by high-performance liquid chromatography using automated pre-column derivatization with o-phthaldialdehyde. J Chromatogr B, Biomed Appl 669:177–186.

FERNSTROM JD, Wurtman RJ (1972) Brain serotonin content: physiological regulation by plasma neutral amino acids. Science 178:414–416.

GEURTS DEM, Huys QJM, Ouden den HEM, Cools R (n.d.) Aversive Pavlovian control of instrumental behavior in humans. Journal of Cognitive Neuroscience: In Press. Available at: http://www.mitpressjournals.org/doi/abs/10.1162/jocn_a_00425

GRAEFF FG, Guimarães FS, De Andrade TGCS, Deakin JFW (1996) Role of 5-HT in stress, anxiety, and depression. Pharmacology Biochemistry and Behavior 54:129–141.

GUITART-MASIP M, Huys QJM, Fuentemilla L, Dayan P, Duzel E, Dolan RJ (2012) Go and no-go learning in reward and punishment: Interactions between affect and effect. NeuroImage 62:154–166.

HAMILTON M (1960) A rating scale for depression. Journal of Neurology, Neurosurgery, and Psychiatry 23:56.

HUYS QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P (2011) Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding Rangel A, ed. PLoS Comput Biol 7:e1002028.

HUYS QJM, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP (2012) Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees Maloney LT, ed. PLoS Comput Biol 8:e1002410.

MIYAZAKI K, Miyazaki KW, Doya K (2012) The role of serotonin in the regulation of patience and impulsivity. Mol Neurobiol 45:213–224.

KIRBY KN, Maraković NN (1996) Delay-discounting probabilistic rewards: Rates decrease as amounts increase. Psychonomic Bulletin & Review 3:100–104.

KRANZ GS, Kasper S, Lanzenberger R (2010) Reward and the serotonergic system. Neuroscience 166:1023–1035.

MAN M-S, Mikheenko Y, Braesicke K, Cockcroft G, Roberts AC (2011) Serotonin at the level of the amygdala and orbitofrontal cortex modulates distinct aspects of positive emotion in primates. Int J Neuropsychopharm 15:91–105.

MCNAUGHTON N, Corr PJ (2004) A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. Neuroscience and Biobehavioral Reviews 28:285–305.

NAKAMURA K, Matsumoto M, Hikosaka O (2008) Reward-dependent modulation of neuronal activity in the primate dorsal raphe nucleus. Journal of Neuroscience 28:5331–5343.

OKADA K-I, Nakamura K, Kobayashi Y (2011) A Neural Correlate of Predicted and Actual Reward-Value Information in Monkey Pedunculopontine Tegmental and Dorsal Raphe Nucleus during Saccade Tasks. Neural Plasticity 2011:1–21.

OLIVER J Robinson BJS (2009) Acute tryptophan depletion evokes negative mood in healthy females who have previously experienced concurrent negative mood and tryptophan depletion. Psychopharmacologia 205:227–235.

PATTON JH, Stanford MS, Barratt ES (1995) Factor structure of the barratt impulsiveness scale. J Clin Psychol 51:768–774.

PAVLOV IP (1927) Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. London: Oxford University Press.

REILLY JG, McTavish SF, Young AH (1997) Rapid depletion of plasma tryptophan: a review of studies and experimental methodology. J Psychopharmacol (Oxford) 11:381–392.

ROBINSON OJ, Cools R, Sahakian BJ (2012) Tryptophan depletion disinhibits punishment but not reward prediction: implications for resilience. Psychopharmacologia 219:599–605.

ROBINSON OJ, Sahakian BJ (2009) Acute tryptophan depletion evokes negative mood in healthy females who have previously experienced concurrent negative mood and tryptophan depletion. Psychopharmacologia 205:227–235

ROBINSON OJ, Sahakian BJ (2008) Recurrence in major depressive disorder: a neurocognitive perspective. Psychol Med 38:315–318.

ROGERS RD, Tunbridge EM, Bhagwagar Z, Drevets WC, Sahakian BJ, Carter CS (2003) Tryptophan depletion alters the decision-making of healthy volunteers through altered processing of reward cues. Nature 28:153–162.

SAULIN A, Savli M, Lanzenberger R (2011) Serotonin and molecular neuroimaging in humans using PET. Amino Acids 42:2039–2057.

SCHMAND B, Bakker D, Saan R, Louman J (1991) [The Dutch Reading Test for Adults: a measure of premorbid intelligence level]. Tijdschrift voor Gerontologie en Geriatrie 22:15–19.

SEYMOUR B, Daw ND, Roiser JP, Dayan P, Dolan R (2012) Serotonin selectively modulates reward value in human decision-making. Journal of Neuroscience 32:5833–5842.

SOUBRIE P (1986) Reconciling the role of central serotonin neurons in human and animal behavior. Behav Brain Sci 9:319–335.

SPIELBERGER CD, gorsuch RL, Lushene R, Vagg PR, Jacobs GA (1983) Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consluting Pscyhologists Press.

THORNDIKE EL (1911) Animal Intelligence. New York: Macmillan.

TORRUBIA R, Åvila C, Moltó J, Caseras X (2001) The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. Personality and Individual Differences 31:837–862.

TYE NC, Everitt BJ, Iversen SD (1977) 5-Hydroxytryptamine and punishment. Nature 268:741–743.

WATSON D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 54:1063–1070.

YIN HH, Ostlund SB, Balleine BW (2008) Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. European Journal of Neuroscience 28:1437–1448.

YOUNG SN, Smith SE, Pihl RO, Ervin FR (1985) Tryptophan depletion causes a rapid lowering of mood in normal males. Psychopharmacologia 87:173–177.

3

AVERSIVE PAVLOVIAN CONTROL OF INSTRUMENTAL BEHAVIOUR IN HUMANS

Geurts, D. E. M., Huys, Q. J. M., Ouden, den, H. E. M., & Cools, R.

Journal of Cognitive Neuroscience, 2013, 25(9), 1428–1441.

Adaptive behaviour involves interactions between systems regulating Pavlovian and instrumental control of actions. Here, we present the first investigation of the neural mechanisms underlying aversive Pavlovian-instrumental transfer using fMRI in humans. Recent evidence indicates that these Pavlovian influences on instrumental actions are action-specific: Instrumental approach is invigorated by appetitive Pavlovian cues, but inhibited by aversive Pavlovian cues. Conversely, instrumental withdrawal is inhibited by appetitive Pavlovian cues, but invigorated by aversive Pavlovian cues. We show that BOLD responses in the amygdala and the nucleus accumbens were associated with behavioural inhibition by aversive Pavlovian cues, irrespective of action context. Furthermore, BOLD responses in the ventromedial prefrontal cortex differed between approach and withdrawal actions. Aversive Pavlovian conditioned stimuli modulated connectivity between the ventromedial prefrontal cortex and the caudate nucleus. These results show that action-specific aversive control of instrumental behaviour involves the modulation of fronto-striatal interactions by Pavlovian conditioned stimuli.

3.1 INTRODUCTION

Adaptive behaviour depends on interactions between systems regulating affective versus rational, instrumental control (Evans, 2008; Huys et al., 2011; Daw et al., 2005). Many decision-making phenomena that appear irrational, such as the framing effect (Tversky & Kahneman, 1981) and the optimism bias (Weinstein, 1980; Sharot et al., 2007), may reflect Pavlovian impact of affective cues on instrumental behaviour (Dayan & Huys, 2008; Dayan et al., 2006). Elucidating the neural mechanisms underlying Pavlovian effects on instrumental actions is crucial, not just for understanding normal behaviour, but also because Pavlovian effects are implicated in neuropsychiatric disorders (e.g. addiction and depression, Dayan & Huys 2008; Flagel et al. 2011). Here we investigate these mechanisms by using fMRI and a well-established paradigm for assessing Pavlovian influences on instrumental responding: Pavlovian-instrumental transfer (PIT).

Existing neuroimaging work on PIT has focused on the potentiation of appetitive instrumental responding by appetitive cues (Talmi et al., 2008; Bray et al., 2008). For example, Talmi et al. (2008) have revealed BOLD responses in the nucleus accumbens and the amygdala during appetitive PIT. However, no imaging study and only a few behavioural studies have addressed the effects of aversive cues on human behaviour (Di Giusto et al., 1974; Huys et al., 2011). This is pertinent, because the influence of aversive expectations on behaviour likely plays an important role in several psychiatric conditions (Bijttebier et al., 2009).

We adapted a paradigm that previously showed significant behavioural PIT of both appetitive and aversive cues (Huys et al., 2011). Our first question was whether structures identified as contributing to appetitive PIT — amygdala and nucleus accumbens — are also involved in aversive PIT. The second question concerned action-specificity, an aspect of PIT that so far has received little attention. We have recently discovered that the effect of Pavlovian cues depended on the valence of instrumental behaviours: While appetitive Pavlovian conditioned stimuli (CSs) potentiated approach and inhibited withdrawal, aversive CSs suppressed approach (as in conditioned suppression), but potentiated withdrawal (Huys et al., 2011). This finding resonates with the fact that many neuropsychiatric disorders prominently involve abnormal control, not only of appetitive behaviours (e.g. approach), but also of aversive behaviours (e.g. withdrawal)(Trew, 2011). If Pavlovian cues have opposite effects on these different actions, then a better understanding of the mechanisms underlying this action-specificity should help resolve how instrumental behaviour is controlled by Pavlovian cues.

Action-specificity suggests that affective cues might interact differently with systems that code for approach or withdrawal. We asked whether action-specificity in Pavlovian control involves differential influences on neural regions that encode action-specificity. One possibility is that it involves direct Pavlovian modulation of regions that encode action-specificity. Another possibility is that Pavlovian cues modulate the influence of regions that are action-specific on regions that implement instrumental behaviour. One region prominently associated with instrumental behaviour is the striatum (the caudate nucleus and putamen) (Balleine & O'Doherty, 2010). We tested these hypotheses by conducting univariate analyses of action-specific PIT effects as well as functional connectivity analyses of action-specific influences on the striatum during PIT.

3.2 MATERIALS AND METHODS

3.2.1 *Subjects*

Fifteen right-handed volunteers participated in a behavioural experiment conducted in a dummy scanner environment prior to the fMRI experiment ("behavioural group"). Subsequently, twenty right-handed volunteers participated in the fMRI experiment ("fMRI group"). The experiment was approved by the local ethics committee. Exclusion criteria were claustrophobia, neurological or cardiovascular diseases, psychiatric disorders, regular use of medication, use of psychotropic drugs, smoking, or metal parts in the body. Written informed consent was obtained before study procedures. Two fMRI subjects were removed from analyses because of below-chance performance in the final stage of the instrumental learning phase and/or during the Pavlovian query trials. For two other fMRI subjects, one of the two sessions was excluded: one subject did not complete the first session due to discomfort in the scanner, while the juice delivery setup failed for another subject's first session. Accordingly, data are reported from 15 subjects (6 women; mean age =25.7; SD =3,4) in the behavioural group and 18 subjects (11 women; mean age = 23.8; SD = 3.5) in the fMRI group.

3.2.2 *Pavlovian-instrumental transfer paradigm*

Subjects performed the experimental task adapted from Huys et al. (2011). The paradigm was programmed using Matlab (2009b, TheMathWorks, Natick, MA) with the Psychophysical Toolbox extension (Brainard, 1997). The experiment consisted of two sessions, each with three stages: (i) instru-

mental training, (ii) Pavlovian conditioning and (iii) Pavlovian-instrumental transfer. The setup of the experiment was the same for the two sessions, but different instrumental and Pavlovian stimuli were used in each session.

Two major adaptations were made to the version used by Huys et al. (2011): First, unlike Huys et al., primary outcomes (juices) were used for Pavlovian conditioning, while secondary outcomes (monetary) were used for instrumental training. This was done to make sure that the (de)motivating effects of the Pavlovian CSs were not due to similarity in outcome with the instrumental action. This made our paradigm sensitive to general as opposed to outcome-specific motivating effects of Pavlovian CSs. Second, subjects had to press a button multiple times rather than just once. This generated an additional dependent variable (the number of button presses), which we anticipated to be sensitive to PIT (cf. Talmi et al., 2008) and allowed us to look at parametric PIT effects in our fMRI analysis.

Instrumental training

The instrumental task (Figure 3.1a) was framed in terms of an approach/withdrawal go/nogo task. On each trial, an instrumental stimulus (mushroom or shell) was presented centrally at the top of the screen. A dot appeared at the bottom of the screen and moved upwards at a constant speed (reaching the top in 2.5s). Subjects had to choose whether to collect the instrumental stimulus by steering the dot through it, or whether not to collect it by steering it past the stimulus. Each choice resulted in monetary wins or losses (+/- 5 cents). Subjects influenced the trajectory of the dot by pressing one button repeatedly. Every button press added a fixed sideways displacement to the dot trajectory. This displacement decayed back to zero over time at a speed that was calibrated before the experimental session to the maximum frequency at which subjects were able to press the button (mean maximum frequency was 4.9 Hz [SD 2.0]). There were two action contexts: In the approach context, the dot appeared in one of the bottom corners and, in the absence of button presses, moved past the instrumental stimulus. Thus, subjects had to actively press the button repeatedly in order to move the dot centrally towards the instrumental stimulus and collect it. In the withdrawal context, the dot appeared in the middle of the screen and by default moved upwards through the instrumental stimulus. In this case, button presses were required to move the dot away from the instrumental stimulus in order to avoid collecting it. Thus there were four trial types: approach-go, approach-nogo, withdrawal-go and withdrawal-nogo (Figure

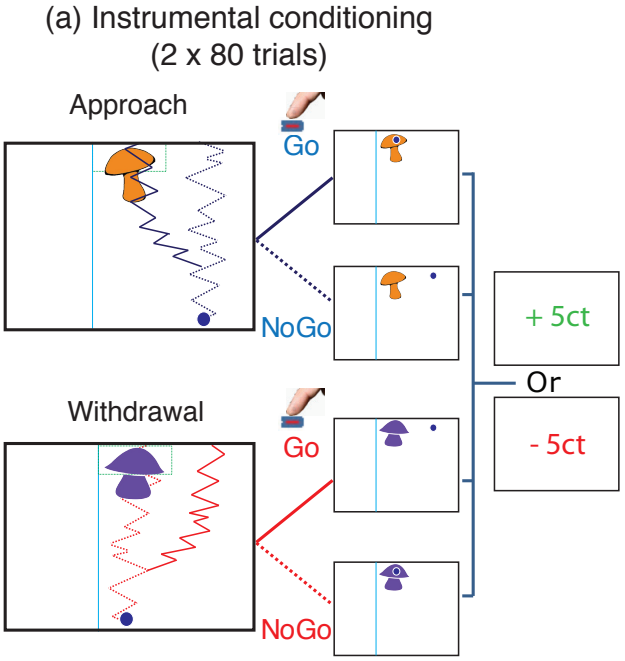

(a) Instrumental conditioning
(2 x 80 trials)
Approach
Go
NoGo
Withdrawal
Go
NoGo
+ 5ct
Or
- 5ct
(b) Pavlovian-instrumental transfer
(2 x 96 trials)
Approach
Go
NoGo
Withdrawal
Go
NoGo
Balance
updated

Figure 3.1 *– Task description.* ***(a) Instrumental training.*** *Trials started with the appearance of the instrumental stimulus at the top center of the screen and of a dot at the bottom of the screen. In approach trials, the dot started either on the left or on the right bottom side of the screen. Subjects could choose to do nothing (approach-nogo), in which case the dot would wiggle past the instrumental stimulus. Alternatively, they could push the button repeatedly to steer the dot through the instrumental stimulus (approach-go). In withdrawal trials, the dot started centrally at the bottom beneath the instrumental stimulus. Subjects could choose to push the button repeatedly to avoid moving through instrumental stimulus (withdrawal-go) or to do nothing (withdrawal-nogo). The four possible trajectories are drawn in the figure (red and blue lines). The green square around the stimulus (invisible to the subject) was the goal region. If the dot entered the goal region, then the instrumental stimulus was collected. The straight line just to one side of the instrumental stimulus was a reflecting boundary that the dot could not cross. Timings were as follows: Instrumental stimuli were presented for 2.5s, during which responses were collected. After 2.5s, feedback was presented for 1s. The inter-trial interval was 1s (blank screen).* ***(b) Pavlovian-instrumental transfer.*** *This paralleled the instrumental training, except that Pavlovian stimuli tiled the background. No outcomes were presented, but subjects were instructed that their choices counted towards the final total. Subjects were explicitly instructed that the juices were collected outside the scanner and they agreed before the start of the experiment to drink them afterwards. Timing of one trial was as follows: 250ms after the onset of the Pavlovian stimulus, the instrumental stimulus (and dot) was overlaid on top of this Pavlovian stimulus. Duration of the instrumental stimulus was 2.5s; duration of the Pavlovian stimulus was 2.75s. Upon offset of both stimuli, feedback was presented, which consisted only of the words "Balance is updated" (duration 1s, ITI 1s).*

3.1a). Thus the Action Context determined whether the active response was an approach or a withdrawal response. Whether an instrumental stimulus was collected was determined based on whether the dot entered a goal region (invisible to the subject; Figure 3.1a) around the instrumental stimulus. If the dot entered this region (after go-approach or nogo-withdrawal), then the stimulus was collected. If not (after nogo-approach or go-withdrawal), then the stimulus was not collected. At times, the dot could touch the target area on the side, only entering it partially. In this case, feedback consisted of the words: "pressed, but incomplete action" and no money was won or lost. At the end of each full action, monetary feedback ("+5 cents" or "-5 cents") was displayed.

In order to orthogonalise the approach-withdrawal and appetitive-aversive axes, the learned instrumental values in approach and withdrawal blocks needed to be matched. To achieve this, both go and nogo responses were, if correct, rewarded to the same extent. Additionally, to avoid a confound of behavioural activation, in each condition (i.e. in both approach and withdrawal conditions) the go action was designated as the correct response for half of the instrumental stimuli, and the nogo action for the other half. Incorrect responses had opposite outcome contingencies to correct responses, yielding more punishments than rewards. This ensured that go, nogo, approach and withdrawal overall had the same learned association with rewards and punishments. In both the approach and withdrawal context, there were 2 go-stimuli, which yielded reward more often after active responses (and punishment after not responding), and 2 nogo-stimuli, which yielded reward more often after not responding (and punishment after go-responding). Reinforcement was probabilistic with probabilities ranging from 0.6 to 1 (on average the ratio reward:punishment following a correct action was 0.85:0.15 for go-stimuli and 0.8:0.2 for nogo-stimuli. The difference arose from a technical error). Trials were labeled as correct if subjects chose the usually rewarded response.

Average reinforcement was matched between approach and withdrawal contexts (behavioural group: mean proportion of positively reinforced trials for approach = 0.58; for withdrawal =0.61, paired sample T-test: T_{14}=-.8, P =.4; fMRI group: mean proportion of positively reinforced trials for approach = 0.63; for withdrawal =0.64; paired sample T-test: T_{17}=-.14, P =.9). Accordingly, the difference between approach and withdrawal actions cannot be driven by Pavlovian responses to the instrumental stimuli. Thus, rather than representing effects of competing Pavlovian responses, the effects we report represent Pavlovian-instrumental transfer effects.

Every session consisted of 80 instrumental training trials alternating between blocks of 8 approach and 8 withdrawal trials. Initial stimuli and action context were randomized across subjects.

Pavlovian conditioning

Each Pavlovian conditioning trial started with the presentation of one of three audiovisual stimuli consisting of a pure tone and a fractal. The appetitive and aversive Pavlovian CSs were followed, respectively, by 2ml of appetitive or aversive juice (i.e. the unconditioned stimuli [US]) on 50% of trials. The neutral CS was followed by no (juice) outcome. Prior to the fMRI experiment, subjects indicated their preference for apple juice, orange juice or strawberry lemonade. The aversive juice was a bitter tasting solution of magnesium sulphate (0.3M). Each Pavlovian CS was presented 20 times, and for each session there was a separate set of 3 stimuli. Stimulus presentation order was fully randomized across subjects. Stimulus duration was 4.5s and juice delivery occurred between 0.5s and 1.5s after stimulus onset. The inter trial interval (ITI) was 1s.

To test and stimulate task involvement during conditioning, query trials were presented after every 10 Pavlovian trials. On these trials, subjects chose one of the two presented Pavlovian stimuli (presented for 2s; ITI 0.5s) in extinction, i.e. there were no outcomes in these trials. The outcomes were only recorded for the last session (due to technical error). In the fMRI group we further assessed conditioning by asking subjects to indicate the degree to which they liked each of the juices and the Pavlovian CSs by means of visual analogue scales (VAS), before and after the experiment.

Pavlovian-instrumental transfer

Stimulus presentation was the same as in the instrumental training stage, except that (i) Pavlovian stimuli tiled the background from 250ms before and during the instrumental trial, and (ii) no monetary feedback and no juice outcomes were presented (Figure 3.1b). However, subjects were instructed that their choices counted towards the final monetary total, and that the juices associated with the Pavlovian stimuli were collected outside the scanner for them to drink afterwards, i.e. PIT was conducted in nominal extinction.

Subjects performed 96 PIT trials per session, alternating between mini-blocks of 8 approach and 8 withdrawal trials. Initial instrumental stimulus, CS and action context were randomized. The numbers of go and nogo stimuli were matched between conditions (i.e. Action Context x CS Valence).

After every trial, feedback consisted of a screen displaying "Balance is being updated".

Image acquisition

Whole-brain imaging was performed on a 3 Tesla MR scanner (Magnetrom Trio Tim, Siemens Medical Systems, Erlangen, Germany). Functional data were obtained using a multi-echo gradient T2*-weighted echo-planar (ME-EPI) scanning sequence (Poser et al., 2006) with blood oxygen level dependent (BOLD) contrast (38 axial-oblique slices, repetition time, 2.32s; echo-times: 9.0, 19.3, 30 and 40ms; in plane resolution, 3.3x3.3 mm; slice thickness, 2.5mm; distance factor 0.17; flip angle, 90). Visual stimuli were projected on a screen and were viewed through a mirror attached to the head coil. In addition, a high-resolution T1-weighted magnetization-prepared rapid-acquisition gradient echo anatomical scan was obtained from each subject (192 sagittal slices; repetition time, 2.3s; echo time, 3.03ms; voxel size 1.0 x 1.0 x 1.0 mm; field of view 256 mm).

3.2.3 *Behavioural data analysis*

The behavioural data were analysed using the statistic software SPSS 16.0 and the modeling was performed in Matlab (2009b, TheMathWorks, Natick, MA).

Instrumental training

First, we assessed change in performance over time during instrumental training. The proportion of correct responses was calculated for the first eight and last eight trials separately for each of the 4 trial types. To assess whether subjects learned to make the correct choice during instrumental training, data were averaged across sessions and submitted to a repeated measures analysis of variance (ANOVA) with Time Bin (2 levels: beginning/end of instrumental training), Action Context (2 levels: approach/withdrawal) and Response Type (2 levels: go/nogo) as within-subject factors and Group (2 levels: behavioural/fMRI) as between-subject factor. Second, we assessed whether the learned behaviour generalized to and over the PIT stage. This was done with the same ANOVA with the difference that the factor Time Bin was changed to include 3 levels: the end of the instrumental training and the beginning and the end of the PIT stage.

Pavlovian conditioning

To assess Pavlovian conditioning, we investigated whether the proportion of correct choices on query trials differed from chance. In addition, liking ratings of the CSs before and after conditioning were analysed using an ANOVA with Time of Rating (2 levels: before/after conditioning) and Valence (3 levels: appetitive/neutral/aversive) as within-subject factors.

Pavlovian-instrumental transfer

There were two dependent measures: choice (go/nogo) and the number of button presses on go-trials. Go-trials were defined as those PIT trials on which one or more than one button press was made. All behavioural outcome measures were averaged across sessions and submitted to ANOVAs with Action Context (2 levels: approach /withdrawal), and Pavlovian CS Valence (3 levels: appetitive/neutral/aversive) as within-subject factors and Group (2 levels: behavioural/fMRI) as between-subject factor. Planned contrasts were targeted at effects of aversive PIT, i.e. the primary focus of this study. For these follow-up analyses, the three-level factor Pavlovian CS Valence in the omnibus ANOVA was replaced by a Pavlovian CS Valence factor with 2 levels: aversive and neutral.

Model-based analyses

We anticipated that the expectation associated with each instrumental stimulus would contribute to the BOLD response. Therefore we computed these expectations (so-called instrumental Q-values) using a reinforcement learning model and included them in the fMRI analysis. The reinforcement learning model and the fitting procedures are described in detail in Huys et al. (2011). After fitting the parameters, the action values Qt1(at; st) determining choice probabilities on trial t were extracted and used in the fMRI analysis.

3.2.4 *fMRI analysis*

fMRI data analysis was performed with SPM5 software (Statistical Parametric Mapping; Wellcome Trust Centre for Cognitive Neuroimaging, London, UK). The first five volumes of each participant's dataset were discarded to allow T1 equilibrium.

First, realignment parameters were estimated for the images acquired at the first echotime and consequently applied to images resulting from the three other echoes. The echo-images were combined by applying a PAID-

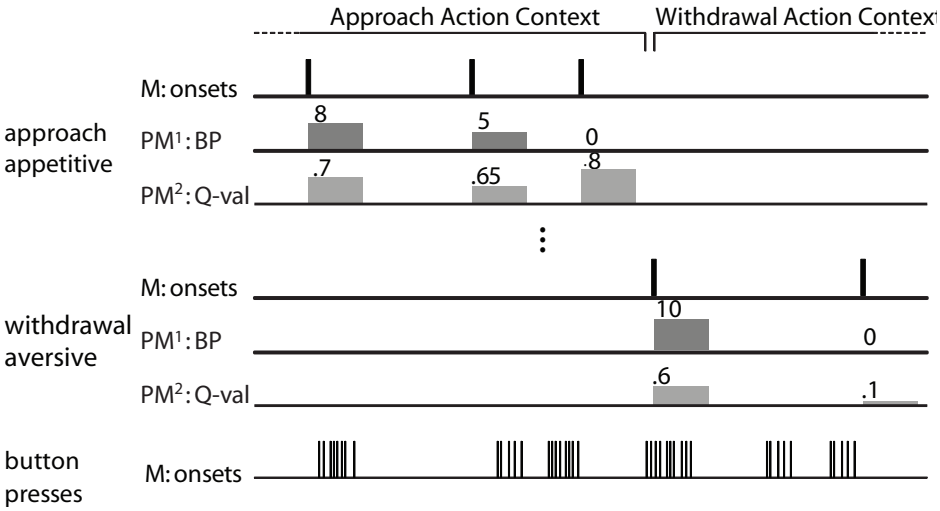


Figure 3.2 *– Schematic depiction of the general linear model to analyse the Pavlovian-instrumental transfer (PIT) data (figure after Talmi et al. (2008)). The main regressors (M) model the onset of a trial as a delta function. There is a main regressor for each of the six trial types. For all six main regressors there are two parametric modulators (PM). The first parametric modulater (PM^1), the PIT-regressor, consists of the number of button presses made per trial (0 for nogo). The second parametric modulator (PM^2) represents the Q-value for each chosen action dependent on the instrumental stimulus shown in the trial at hand. In the 7th main-regressor (of no interest) every single button press is modeled by a delta function. For reasons of clarity, two of the six trial types (approach appetitive and withdrawal aversive) are depicted only for one session and no movement nuisance regressors are shown.*

weight algorithm assessing the signal-to-noise ratio as described by Poser et al. (2006). Thirty volumes, acquired before each instrumental training session, were used as input for this algorithm. Thereafter the following preprocessing steps were applied: slice-time correction, coregistration and a segmentation procedure using the tissue probability maps provided by SPM5 for grey matter, white matter and CSF centered in MNI space to estimate normalization parameters based on the structural image. Structural as well as functional images were then normalized by applying these estimations. All normalized images were smoothed with an isotropic 8 mm full-width half-maximum Gaussian kernel (Worsley & Friston, 1995).

A random effects, event-related, statistical analysis was performed with SPM5. This analysis was restricted to the PIT-stage. First, we specified

a separate general linear model (GLM) for each participant (Figure 3.2). For each session six main regressors represented the six different PIT trials: (1) approach appetitive, (2) approach neutral, (3) approach aversive, (4) withdrawal appetitive, (5) withdrawal neutral and (6) withdrawal aversive. For each main regressor two additional parametric regressors were added (Büchel et al., 1996): (i) One regressor represented the tonic parametric modulation of BOLD responses during each trial by the number of button presses per trial; the PIT-regressor (cf. Talmi et al., 2008). (ii) Another regressor represented the parametric modulation of BOLD responses by the Q-value per trial as estimated from the model-based analysis. These parametric modulators were serially orthogonalized. Additionally, a regressor of no interest modeled phasic button presses as single events (cf. Talmi et al., 2008). All paradigm-related regressors were modeled as delta functions at the onset of the instrumental stimulus presentation per trial and were convolved with a canonical hemodynamic response function (HRF). Realignment parameters (the three rigid-body translations and three rotations) were added to capture residual movement-related artifacts. High-pass filtering (128s) was applied to the time series of the functional images to remove low-frequency drifts. Parameter estimates for all regressors were obtained by maximum-likelihood estimation, modeling temporal autocorrelation (AR1). The parameter estimates, derived from this fit of the model to the data, reflect the strength of covariance between the data and the canonical response function for each of the regressors.

Parameter estimates for the six parametric PIT-regressors were estimated at the subject-level and then used in a 2x3 ANOVA (full factorial design) at the group-level with factors Action Context (2 levels: approach/withdrawal) and CS Valence (3 levels: appetitive/neutral/aversive) as within-subject factors. Restricted Maximum Likelihood estimates of variance components were used to allow for unequal variance between subjects and possible deviations from sphericity introduced by dependencies between levels in the repeated measures design. The main effects and interactions were then calculated. We assessed the following three, planned contrasts to test our hypotheses, which focused on aversive PIT:

1. Main effect of CS Valence, contrasting aversive and neutral CSs ([approach neutral + withdrawal neutral] - [approach aversive + withdrawal aversive]). This contrast identified CS-dependent coupling of the BOLD response with the number of button presses, i.e. aversive PIT-related BOLD responses independent of Action Context.

2. Interaction between Action Context and CS Valence ([approach neutral – approach aversive] - [withdrawal neutral - withdrawal aversive]). This contrast identified BOLD responses associated with action-specific aversive PIT.

3. Main effect of Action Context, contrasting approach and withdrawal trials ([approach appetitive + approach neutral + approach aversive] - [withdrawal appetitive + withdrawal neutral + withdrawal aversive]). This contrast identified regions where BOLD responses are action specific, i.e. differ between approach and withdrawal.

To investigate the valence-specificity of the effects, supplementary analyses were conducted to assess the same three contrasts, with the CS Valence factor contrasting appetitive with aversive CSs, and appetitive with neutral CSs.

It is important to note that the parametric nature of the PIT-regressor ensures that the contrasts of interest represent BOLD response involved in PIT, and do not reflect differences in motor activity or Pavlovian CS per se (cf. Talmi et al., 2008). However, this analysis explicitly discounts signals that are constant, i.e. do not vary as a function the number of button presses during the presentation of each CS. Therefore, following Talmi et al (2008), to take such signals into account we also contrasted the main regressors (instead of the parametric PIT-regressors) at the subject-level to calculate both a main effect of CS Valence ([approach neutral + withdrawal neutral] - [approach aversive + withdrawal aversive]) and an interaction between CS Valence and Action Context ([approach neutral - approach aversive] - [withdrawal neutral - withdrawal aversive]). The resulting statistical parametric maps for each contrast were then used to conduct a t-test at the group-level with behavioural aversive PIT-effects as a covariate. The behavioural aversive PIT effect for each subject was computed in terms of the average number of button presses, irrespective of Action Context ([approach neutral + withdrawal neutral] - [approach aversive + withdrawal aversive]), and as a function of Action Context ([approach neutral - approach aversive] - [withdrawal neutral - withdrawal aversive]). These analyses revealed regions in which CS-dependent BOLD responses were associated with individual behavioural PIT effects.

Functional connectivity analyses

Model-based analyses Next we assessed whether action-specificity of behavioural aversive PIT was accompanied by action-specific PIT-related functional connectivity. Specifically, we conducted psychophysiological interac-

tion (PPI) analysis to assess whether action-specific PIT was associated with PIT-related modulation of functional connectivity with seed regions exhibiting a main effect of Action Context. First, for each individual the (first principal component of the) BOLD time series was extracted from an 8 mm sphere surrounding the BOLD response peak revealed by the main Action Context contrast (the seed) (the ventromedial prefrontal cortex; Figure 3.6). The time series was then deconvolved based on the canonical hemodynamic response model to construct a time series of neural BOLD responses following the procedures outlined by Gitelman et al. (2003). Second, for every subject two GLMs were estimated, one for each Action Context, which included the following three regressors (as well as the six motion parameters): (1) The seed BOLD response time series; (2) a parametric task contrast regressor representing aversive PIT (neutral minus aversive); and (3) the PPI regressor, i.e. the interaction between (1) and (2), computed by multiplication of the deconvoluted regressor (1) and regressor (2). The PPI-regressor was then convolved with the HRF. Parameter estimates for the PPI-regressor were estimated by maximum-likelihood estimation, modeling temporal autocorrelation (AR1) at the subject-level, and were then used in a t-test at the group-level. The parameter estimates, derived from this fit of the model to the data, reflect the strength of PIT-related connectivity with the action-specific seed region (the ventromedial prefrontal cortex). To assess the relationship between individual behavioural PIT-effects and functional PIT-related connectivity, covariates representing behavioural PIT effects (average number of button presses during neutral minus aversive trials) were included in the second level group analysis.

Statistical thresholding and volumes of interest

We report only those effects that survive family wise error (FWE) correction for multiple comparisons at the whole brain (PFWE WB <.05, voxel-level) or within volumes of interest (PFWE SV <.05, voxel-level). Based on existing literature (Talmi et al., 2008; Corbit, 2005; Corbit & Balleine, 2011), we expected PIT effects in the amygdala and the nucleus accumbens. Therefore, these regions were defined as volumes of interest, using anatomical criteria. The bilateral amygdala was defined using the automated anatomical labelling atlas (Tzourio-Mazoyer et al., 2002). The bilateral nucleus accumbens was segmented for each subject using the FSL FIRST segmentation tool (Patenaude et al., 2011). These individual segments were then overlaid onto each other, generating one nucleus accumbens for the group. The amygdala and accumbens volumes were combined, so that voxel-level correction for multiple comparisons was conducted for all voxels within

these two volumes. Furthermore, we had a specific hypothesis regarding the action-specificity of the PIT effects. In particular, we reasoned that action-specificity of PIT might arise from Pavlovian effects on neural regions known to implement instrumental action. One of the most prominent regions implicated in instrumental action control is the striatum (Balleine & O'Doherty, 2010). Therefore, we conducted additional (univariate and connectivity) analyses of action-specific effects in the bilateral striatum, defined as the caudate nucleus and putamen based on the automated anatomical labeling atlas (Tzourio-Mazoyer et al., 2002).

3.3 RESULTS

3.3.1 *Behavioural data*

Behavioural data are reported across the behavioural (n=15) and fMRI group (n=18). To facilitate interpretation of the fMRI results, we additionally present the data for the fMRI group separately. However, no significant differences between the groups were found.

Instrumental conditioning

Analysis of the first stage of the experiment indicates robust instrumental learning (Figure 3.3a). Subjects learned to make correct choices during the instrumental learning stage indicated by an increasing number of correct responses over time ($F_{1;31}$=97,9 , P <.001). Furthermore, these differences were affected by the action context (approach or withdrawal) and changed over time: There was a significant three-way interaction between Time Bin, Action Context and Response Type ($F_{1;31}$=34.0, P <.001). This was due to subjects initially preferring to approach the instrumental stimulus, i.e. to go during approach (approach-go vs nogo at the beginning of instrumental training: T_{32}=8.9, P <.001) but to nogo during withdrawal (withdrawal-go vs nogo: T_{32}=-4.6, P <.001; simple interaction effect between Action Context and Response Type at the end of instrumental training: $F_{1;31}$=87.6, P <.001). The bias towards withdrawal-nogo disappeared and the bias towards approach-go became less strong, but remained significant during learning (withdrawal-go vs nogo at the end of instrumental training: T_{32}=-1.4, P >.1; appoach-go vs nogo at the end of instrumental training: T_{32}=2.4, P <.05; simple interaction effect between Action Context x Response Type at the end of instrumental training: $F_{1;31}$=7.1, P <.05).

In addition, as is also explained by the interactions described in the previous paragraph, there was a significant interaction between Time Bin and Action Context ($F_{1;31}$=5.7, P <.05) and a significant interaction between Response Type and Action Context across Time Bins ($F_{1;31}$=58.6, P <.001). Furthermore, there was a main effect of Response Type, due to subjects making more correct go responses than correct nogo responses across the instrumental training (main effect of Response Type: $F_{1;31}$=13.4, P <.01). There was no significant main effect of or interaction with the factor Group.

For the fMRI group alone almost the same pattern was found as for the whole group: Subjects learned to make correct choices during the instrumental learning stage indicated by an increasing number of correct responses over time ($F_{1;17}$=50.0 , P <.001; Figure 3.3b). Additionally there was a three-way interaction between Time Bin, Action Context and Response Type ($F_{1;17}$=26.2, P <.001). Again this was driven by the initial inclination of subjects to collect the instrumental stimulus: initially subjects preferred to go during approach (paired sample T-test: T_{17}=5.3, P <.001) but to nogo during withdrawal (paired sample T-test: T_{17}=-2.4, P <.05; simple interaction effect between Action Context and Response Type at Time Bin 1: $F_{1;17}$=35.6, P <.001; Figure 3.3b). These biases were overcome at the end of the instrumental training stage (paired sample T-test: T_{17}=0.5, P >.1; T_{17}=-0.2, P >.1; simple interaction effect between Action Context x Response Type at Time Bin 2: $F_{1;17}$=.2, P >.1). In addition there was a significant interaction between Response Type and Action Context across Time Bins ($F_{1;17}$=10.9, P <.01).

Instrumental generalization to the PIT stage

Performance at the end of instrumental training generalized to, and persisted throughout the PIT stage (Figure 3.3a): there were no significant main effects of, or interactions with Time Bin when the 2-level factor Time Bin was replaced with a Time Bin factor with 3 levels: the end of the instrumental training, the beginning of the PIT stage and the end of the PIT stage. This was also the case when considering data from the fMRI group only (Figure 3.3b).

Pavlovian conditioning

In both groups, analysis of the Pavlovian query trials confirmed successful Pavlovian conditioning (mean proportion correct in behavioural group: 95%; SEM: 3.1; range: 58-100%; fMRI group: 94%; SEM: 1.9; range: 80-100%).

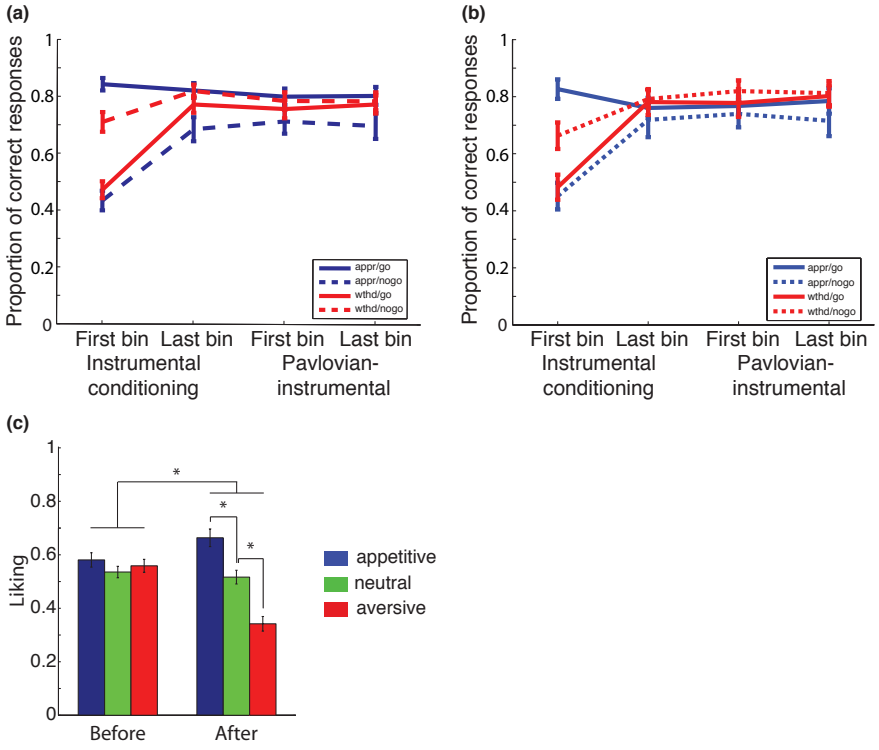


Figure 3.3 *– Instrumental learning and generalization to the Pavlovian-instrumental transfer stage for* ***(a)*** *the whole group and for* ***(b)*** *the fMRI group separately. The proportion of correct choices are broken down by Response Type (go/nogo) and Action Context (approach/withdrawal). Error bars represent standard errors of the mean.* ***(c)*** *Visual analogue scale ratings before and after Pavlovian conditioning. Bars represent group means of visual analogue scale scores (0=very aversive, 0.5 = neutral, 1=very appetitive). Error bars represent standard errors of the mean (* = P<.05).*

A one-sample T-test on the liking ratings of the US (i.e. juices; only available for the fMRI group) showed that, at baseline (pre), subjects judged the aversive US to be aversive (mean pre=.21, significantly different from 0.5: T_{17}=14.4, P <.001, [scores ranged from 0 (aversive) to 1 (appetitive) with 0.5 indicating neutral]) and the appetitive US to be appetitive (significantly different from 0.5: mean pre=.70, T17=4.1, P =.001). Ratings for the aversive US did not change significantly over the course of the experiment (pre versus post, paired sample T-test: meanpost=.20, T_{17}=.2, P >.05); the appetitive US became slightly more appetitive across time (paired sample T-test: meanpost=.80, T_{17}=2.7, P <.05) (CS Valence x Time: $F_{1;17}$=5.1, P <.05).

VAS ratings for the Pavlovian CSs (only available for the fMRI group) showed that Pavlovian conditioning induced changes in subjective liking (ANOVA Time x CS Valence: $F_{1.3;22.3}$=10.6, P =.002; Figure 3.3c). Simple Time (pre/post) x CS Valence (2 levels) interaction analyses confirmed that conditioning altered ratings for the aversive relative to the neutral CS ($F_{1;17}$=9.6, P =.007), for the appetitive relative to the neutral CS ($F_{1;17}$=6,0, P =.026) and for the appetitive relative to the aversive CS ($F_{1;17}$=12.4, P =.007). There were no differences between the three CSs prior to conditioning (paired sample T-test: appetitive versus neutral (T_{17}=1.5, P >.1), appetitive versus aversive (T_{17}=.8, P >.1), neutral versus aversive (T_{17}=-.6, P >.1)). Conversely, after conditioning, liking ratings were significantly higher for the neutral than for the aversive CS ($F_{1;17}$=10.9, P <.01), for the appetitive than for the neutral CS ($F_{1;17}$=9,7, P <.01) and for the appetitive than for the aversive CS ($F_{1;17}$=24,5, P <.001).

Pavlovian-instrumental transfer

Analysis of choice (go vs nogo) data from the PIT stage revealed a significant action-specific PIT effect, which partially replicated that reported by Huys et al. (2011). Thus the proportion of approach-go responses was lower during display of the aversive CS than that during display of the neutral CS (i.e. subjects exhibited conditioned suppression). Conversely, the proportion of withdrawal-go responses was higher during display of the aversive CS than that during display of the neutral CS (Figure 3.4). This observation was confirmed statistically by a significant two-way interaction between Action Context (approach vs withdrawal) and CS Valence (aversive vs neutral) (for the group as a whole: $F_{1;31}$=6.8, P <.05; for the fMRI group only: $F_{1;17}$=3.3, P =.085). Furthermore, simple effects analyses confirmed the presence of statistically significant simple effects of CS Valence (aversive vs neutral) for approach (whole group: $F_{1;31}$=5.4, P <.05; fMRI group only: $F_{1;17}$=2.1, P >.1) as well as for withdrawal (whole group: $F_{1;31}$=5.1, P <.05; fMRI group only:

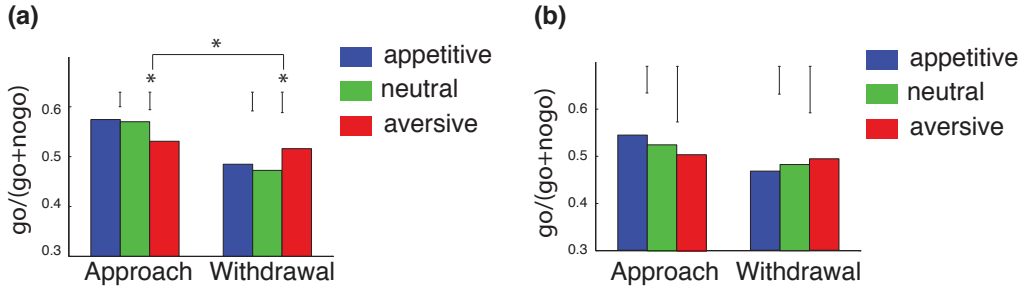


Figure 3.4 *– Behavioural data from the Pavlovian-instrumental transfer stage. Shown are choice data as a function of Action Context (approach and withdrawal) and CS Valence (appetitive/neutral/aversive) for* ***(a)*** *the whole group and* ***(b)*** *the fMRI group separately. Error bars represent standard errors of the difference between, respectively, trials with appetitive and neutral CSs, and trials with aversive and neutral CSs (* = P<.05).*

$F_{1;17}$=.4, P >.1). Thus, our task successfully revealed aversive PIT, an effect that was action-specific.

In contrast, we did not find evidence for appetitive PIT. On the one hand, the omnibus F-test with CS Valence as a three- instead of two-level factor (appetitive vs neutral vs aversive) did reveal a significant two-way interaction between Action Context and CS Valence (whole group: $F_{2;62}$= 4.2, P <.05; fMRI group only: $F_{2;34}$=2.3 P >.1 [linear contrast: $F_{1;17}$=3.7, P =.069]), However, in contrast to our hypotheses, when appetitive CSs were compared with neutral CSs, there was no simple main effect of CS Valence (whole group: for approach: $F_{1;17}$=0.9 P >.1; for withdrawal: $F_{1;17}$=0.4 P >.1), and no simple interaction effect between Action Context and CS Valence (whole group: $F_{1;31}$=0.7, P >.1). This suggests that our task was not appropriate for measuring appetitive PIT.

Irrespective of CS Valence, subjects made more go-responses in the approach than in the withdrawal context (whole group: main effect of Action Context ($F_{1;31}$=4.4, P <.05). This main effect of Action Context concurs with the pattern of performance in the initial instrumental training stage, which also revealed a main effect of Action Context.

There were no significant effects of the factor Group (behavioural/fMRI). Consistent with this lack of effect, the performance patterns were similar when analysed separately for the fMRI group (Figure 3.4b), although the effects did not reach statistical significance (for stats see above).

	Action Context	
	Approach	**Withdrawal**
Appetitive	8.64 (0.27)	8.71 (0.30)
Neutral	8.78 (0.34)	8.66 (0.26)
Aversive	8.33 (0.32)	8.47 (0.36)

Table 3.1 *– Presented are the average number of button presses for the fMRI group as a function of Action Context (approach/withdrawal) and CS Valence (appetitive/neutral/aversive) during the Pavlovian-instrumental transfer stage (standard errors of the mean).*

There were no effects in terms of the total number of button presses (Table 3.1).

3.3.2 *Imaging data*

BOLD responses in the amygdala and nucleus accumbens during aversive PIT

We first performed an ANOVA using the parametric PIT regressors, and with Action Context (approach/withdrawal) and CS Valence (appetitive/neutral/aversive) as within-subject factors. There were no main effects of CS Valence, and no interactions between Action Context and CS Valence, as revealed by whole brain analyses and by small volume analyses (of the amygdala, nucleus accumbens and the striatum).

However, when taking individual differences in behavioural PIT effects into account, we observed significant brain-behaviour correlations in the amygdala and the nucleus accumbens: Subjects who exhibited greater aversive inhibition of instrumental responding (across approach and withdrawal contexts) showed higher BOLD responses during aversive relative to neutral CSs (Figure 3.5). This was revealed by an ANOVA with the main regressors and the behavioural aversive PIT effect in terms of button presses as a covariate.

These brain-behaviour correlations were due to significant associations between individual differences in the behavioural aversive PIT effect and BOLD responses in the bilateral amygdala and in the left nucleus accumbens. These effects in the amygdala and nucleus accumbens were present irrespective of Action Context. These analyses did not reveal any action-specific brain-behaviour correlations, even when analyzed within our small volumes including the striatum. Thus BOLD responses in the amygdala

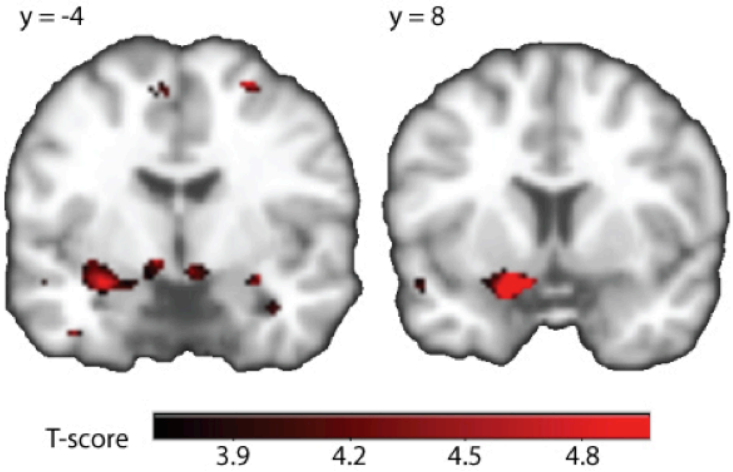


Figure 3.5 *– Aversive PIT related BOLD response in the bilateral amygdala and left nucleus accumbens. The left image depicts regions of the amygdala (bilateral) where change in BOLD response between neutral and aversive CS trials was positively related to behavioural inhibition during aversive CS trials compared to neutral CS trials (Small volume correction with the nucleus accumbens and amygdala volume of interest: T=5.45 , P_{FWE-SV} =.009 , MNI coordinates of peak voxel: xyz =[-30 -4 -16]; T=4.47 , P_{FWE-SV} =.044 , xyz = [32 -4 -14], covariate: mean=0.32, SD=0.71). The right image shows that the same effect is significant for the left nucleus accumbens (T=4.97 , P_{FWE-SV} =.020 , xyz = [-14 8 -14]). Images are displayed at a statistical threshold of P <.001 uncorrected.*

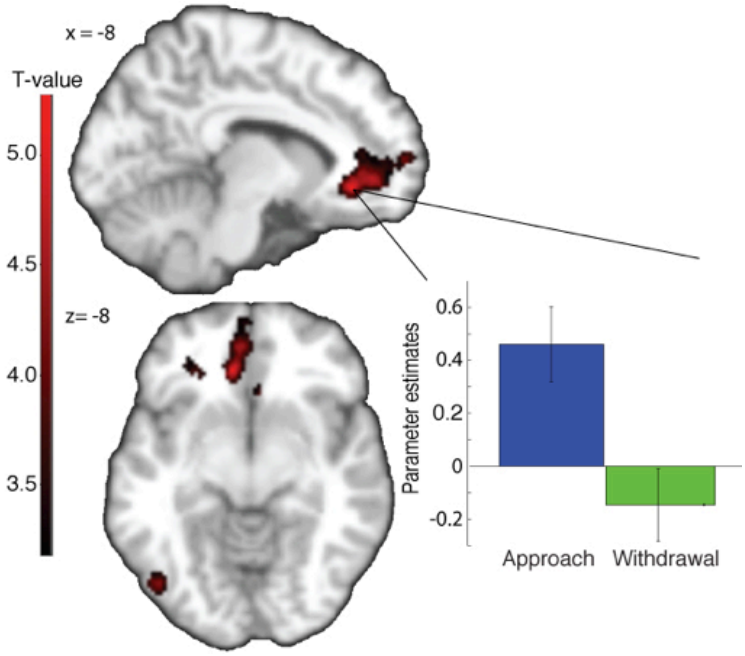


Figure 3.6 *– Action-specific BOLD response in the ventromedial prefrontal cortex. There was a main effect of Action Context in the ventromedial prefrontal cortex (T=5.25, P_{FWE-WB}=.019, MNI coordinates of peak voxel: xyz=[-8 36 -8]). The bar graph shows parameter estimates from the peak voxel for the different Action Contexts. Images are displayed at a statistical threshold of P <.001 uncorrected.*

and nucleus accumbens to aversive CSs predicted individual differences in aversive Pavlovian inhibition, in a manner that was independent of Action Context.

The effects were also unique to aversive CSs and did not extend to appetitive CSs: a supplementary analysis contrasting appetitive and neutral CSs did not yield effects. Furthermore, supplementary analyses comparing aversive and appetitive CSs did not reveal the effects seen above in the comparison between aversive and neutral CSs. This lack of effect when including appetitive CSs might be due to increased variability during the appetitive CSs.

Ventromedial prefrontal cortex differentiates between approach and withdrawal context

Whole brain ANOVA with the parametric PIT regressors (Action Context [approach/withdrawal] X CS Valence [appetitive/neutral/aversive]) revealed a main effect of Action Context in the ventromedial prefrontal cortex (Figure 3.6). BOLD responses in this region were higher in the approach than in the withdrawal context. The inverse effect was observed in the bilateral lingual gyrus (T=9.57, P_{FWE-WB}= .001, MNI coordinates: xyz=[16 -74 0] and [-2 -78 18]) and in the bilateral precuneus (T=6.0, PFWE WB=.001, xyz=[10 -54 48] and [-10 -48 48]). Small volume analyses of responses in the amygdala, nucleus accumbens and striatum did not reveal any subcortical action specificity.

Action-specificity of aversive PIT is accompanied by action-specific fronto-striatal connectivity

Next we assessed whether action-specificity of behavioural aversive PIT was accompanied by PIT-related functional connectivity with this action context-specific BOLD response in the ventromedial prefrontal cortex. To this end, we conducted psychophysiological interaction analyses, separately for the approach and the withdrawal context, with the action-specific ventromedial prefrontal cortex region as the seed (Figure 3.6) and with a task contrast regressor representing aversive PIT (the number of button presses for aversive versus neutral CSs).

When individual differences in behavioural PIT effects were not taken into account, small volume analyses revealed a significant effect in the striatum (centered on the caudate nucleus; MNI coordinates: xyz=[-12 20 4]) for withdrawal, but not approach. Specifically, in the withdrawal condition, there was a significant positive contribution of the ventromedial prefrontal cortex to the caudate nucleus during aversive PIT (Figure 3.7a). Thus, PIT-related connectivity between the ventromedial prefrontal cortex and the caudate nucleus was higher during aversive than during neutral CSs. No such effects, across the group as a whole, were found for the approach condition. However, when individual differences in behavioural PIT effects were taken into account, both small volume and even whole brain analyses revealed a significant effect for the approach condition, again in the caudate nucleus (MNI coordinates: xyz=[-4 12 12] and [-14 26 2]; Figure 3.7b). This effect reflected a negative association between the behavioural aversive PIT effect and the psychophysiological interaction effect: Greater aversive Pavlovian inhibition of approach responding was associated with reduced

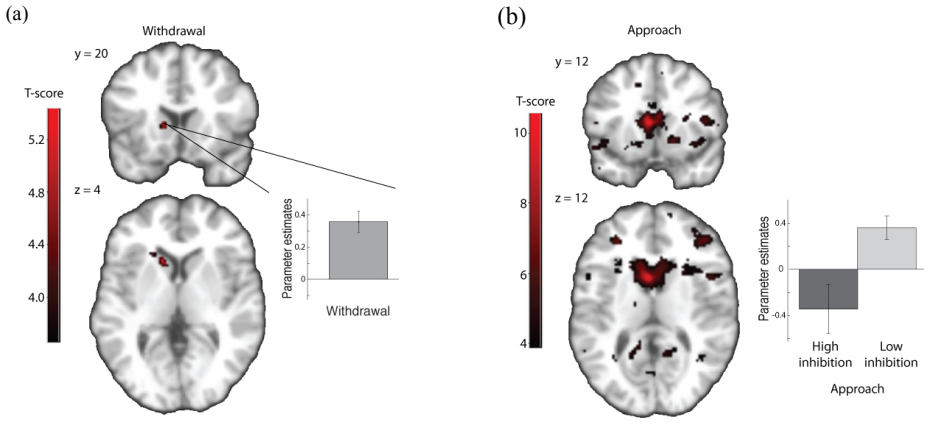


Figure 3.7 *– Functional connectivity during action-specific, aversive Pavlovian-instrumental transfer (PIT). (a) The PPI analysis of aversive PIT in withdrawal showed PIT-related connectivity between the left caudate nucleus and the ventromedial prefrontal cortex (Small volume correction with the striatum volume of interest: T=4.08, P_{FWE-SV} =.031, xyz=[-12 20 4]). The bar graph shows parameter estimates from the peak voxel. This reveals that PIT-related connectivity between the ventromedial prefrontal cortex and the caudate nucleus was higher during aversive than during neutral trials. (b) For aversive PIT in approach the brain image shows that PIT-related connectivity between the ventromedial prefrontal cortex and the caudate nucleus was associated with behavioural PIT-effects (Family wise error correction for multiple comparisons for the whole brain: T=10.50, P_{FWE-WB}=.001, xyz=[-4 12 12]; T=9.75, P_{FWE-WB}=.002, xyz=[-14 26 2], covariate: mean=.45, SD=1.05). To interpret this association, parameter estimates from the peak voxel of the PPI analysis are shown in the bar graph for subjects with high and low behavioural aversive PIT-effects, i.e. with high and low behavioural inhibition during presentation of the aversive CS (median split). This reveals that PIT-related connectivity between the ventromedial prefrontal cortex and the caudate nucleus was lower during aversive than during neutral trials for subjects who showed more behavioural inhibition. Images are displayed at a statistical threshold of P <.001 uncorrected.*

connectivity between the ventromedial prefrontal cortex and the caudate nucleus during aversive relative to neutral CSs. Thus, action-specific signal in the ventromedial prefrontal cortex contributed in a CS-dependent manner to the BOLD signal in the caudate nucleus. The same effect was significant in the bilateral nucleus accumbens (Small volume correction with the nucleus accumbens and amygdala volume of interest: T=4.98, P_{FWE-SV} =.017, MNI coordinates: xyz=[10 18 -2]; T=4.82, P_{FWE-SV} =.017, xyz=[-12 10 -8]). This effect was unique to the striatum and the nucleus accumbens, as whole brain and small volume correction analysis did not reveal any other meaningful effects.

3.4 DISCUSSION

The present study addressed two key questions concerning human Pavlovian-instrumental transfer (PIT). First, unlike prior studies, it revealed the neural mechanisms underlying PIT in the aversive domain and enabled us to conclude that the human amygdala and nucleus accumbens are involved in the effects of aversive Pavlovian cues on instrumental behaviour. Second, this study addressed, for the first time, the neural mechanisms underlying action-specificity of human PIT. Differential responses for approach and withdrawal were found in the ventromedial prefrontal cortex. Furthermore, aversive CSs modulated functional connectivity between the ventromedial prefrontal cortex and the caudate nucleus, both regions strongly associated with goal-directed instrumental control (Balleine & O'Doherty, 2010; Valentin et al., 2007). These results suggest that one origin of action-specificity of PIT lies in the engagement of goal-directed control systems, such as the ventromedial prefrontal cortex and the caudate nucleus, and involves Pavlovian regulation of goal-directed fronto-striatal circuitry.

These findings generally concur with long established observations that the ventromedial prefrontal cortex is key for the affective control of behaviour (Damasio & Everitt, 1996; Damasio, 1997; Greene, 2001; Clark & Manes, 2004; Wallis, 2007; Rushworth et al., 2011). Indeed, this region receives abundant input from regions that process affective information including the amygdala and the nucleus accumbens (Mayberg et al., 1999; Haber, 2003; Haber & Knutson, 2010; Ongür & Price, 2000), and it is critical for the instrumental guidance of behaviour by representations of current goals (Valentin et al., 2007). Furthermore, recent electrophysiological findings in rats suggest that subsets of neurons in the ventromedial prefrontal cortex are involved in the integration of Pavlovian and instrumental infor-

mation that underlies PIT (Homayoun & Moghaddam, 2009). This fMRI study did not reveal PIT signals in the ventromedial prefrontal cortex that evidence such integration. However, the prefrontal cortex is well known not to act alone in guiding decision-making, but interacts with a set of strongly connected subcortical structures via fronto-striatal circuits (Alexander et al., 1986; Haber, 2003; Haber & Knutson, 2010). In keeping with this, we found PIT-related connectivity between the ventromedial prefrontal cortex and the caudate nucleus as well as PIT-related signals in subcortical structures, such as the amygdala and nucleus accumbens.

Our study aimed specifically to address the neural mechanisms of action-specificity in PIT. The finding that the ventromedial prefrontal cortex codes action-specificity was obtained despite the fact that the values of approach and withdrawal goals (or actions) were the same (paired sample T-test on action/Q-values: T_{17}=-1.5, P>.1). This is remarkable given previous work showing an important role for the ventromedial prefrontal cortex in representing goal (or action) values (Kahnt et al., 2011; Wit et al., 2009; Hare et al., 2010; Rangel et al., 2008; Kable & Glimcher, 2009; Hare et al., 2009). Its implication in goal-directed control is substantiated by another previous finding showing that BOLD responses in this region change as a function of outcome devaluation (Valentin et al., 2007). Our finding that approach behaviour engages the ventromedial prefrontal cortex to a greater extent than does withdrawal behaviour might reflect the fact that, in this paradigm, there is an asymmetry between approach and withdrawal. Because the goal state (the instrumental stimulus) is more clearly delineated for approach than for withdrawal, it is conceivable that approach behaviour is driven more readily by a goal-directed system (critically involving the ventromedial prefrontal cortex) than withdrawal behaviour. According to an alternative, not mutually exclusive account, the differential response in the ventromedial prefrontal cortex might also reflect differences in visual attention paid to the goal state. Indeed, Lim et al. (2011) have recently shown that the ventromedial prefrontal cortex encodes (relative) value signals as a function of visual attention. This hypothesis also concurs with the finding that in our paradigm BOLD effects in visual occipital regions differentiated withdrawal from approach. Thus action-specificity in this PIT task might originate in systems that represent action values in a manner that is modulated by the goal state space and/or visual attention.

The observation that aversive PIT was accompanied by Pavlovian modulation of influences from this action-specific ventromedial prefrontal cortex on the caudate nucleus further strengthens the hypothesis that action-specificity in PIT involves modulation of goal-directed control systems. In-

deed the rodent homologue of the caudate nucleus, i.e. the dorsomedial striatum, has also been shown to be sensitive to changes in outcome devaluation (Yin, Ostlund, et al., 2005; Yin, Knowlton, & Balleine, 2005). Furthermore, our findings reveal a strong relationship between the inhibition of instrumental approach by aversive Pavlovian cues and disruption of fronto-striatal connectivity by aversive Pavlovian cues. Based on this result, we speculate that aversive Pavlovian inhibition of approach (i.e. conditioned suppression) is accompanied by frontal suppression of striatal processing. The reverse pattern was observed for withdrawal, in which fronto-striatal connectivity was enhanced by the aversive cues, consistent with the speculation that aversive Pavlovian potentiation of withdrawal is accompanied by frontal enhancement of striatal processing. This proposal generally concurs with ideas that choice and planning of appropriate actions are instantiated by spiralling fronto-striatal pathways, including those connecting the ventromedial prefrontal cortex and the caudate nucleus (Haber et al., 2000; Balleine & O'Doherty, 2010). Our connectivity findings indicate that processing in these pathways can be modulated by aversive Pavlovian CSs. This chimes well with our recent findings that inhibitory Pavlovian responses are able to significantly constrain goal-directed choice behaviour (Huys et al., 2012).

The observation that the amygdala and the nucleus accumbens are involved in PIT concurs with animal studies showing that the influence of appetitive Pavlovian cues on instrumental decision-making depends on the integrity of the amygdala and nucleus accumbens (Corbit & Balleine, 2011; Corbit, 2005). These studies have suggested that the amygdala represents the affective valence of Pavlovian cues, while the nucleus accumbens is thought to represent a limbic-motor interface, transmitting affective information to the spiraling cortico-striatal pathways. Our findings are also consistent with results from a study in humans revealing activity in both these regions during appetitive PIT (Talmi et al., 2008). That study showed that appetitive Pavlovian effects on instrumental vigour were associated with BOLD signal in the ventral striatum during appetitive cues compared with neutral cues. In addition, brain-behaviour associations showed that subjects who exhibited stronger behavioural PIT also exhibited stronger responses in the ventral striatum and amygdala. The key conclusion of the present study is that these regions are also involved in aversive PIT. Unlike the pattern of responses in the ventromedial prefrontal cortex, and unlike the pattern of connectivity with the caudate nucleus, the responses in the amygdala were not action-specific, suggesting that it participates in Pavlovian inhibition of instrumental actions regardless of their approach/withdrawal nature.

The primary interest of this study was to uncover neural mechanisms of aversive rather than appetitive PIT. However, it is notable that, unlike prior work, the present study did not replicate an effect of appetitive PIT in the amygdala or in the nucleus accumbens (cf. Talmi et al., 2008). We emphasize that our failure to demonstrate appetitive PIT does not diminish the validity of the paradigm for measuring aversive PIT. Nevertheless, in the following we consider a few hypotheses regarding this lack of effect. One key difference is that we used different outcomes for the Pavlovian and instrumental training stage and that our paradigm therefore captures exclusively outcome-general PIT. Talmi et al. (2008) used the same outcomes for both stages, and the effects they see in the nucleus accumbens and amygdala could therefore conceivably be driven by both outcome-general and outcome-selective PIT effects. Although animal work does suggest that both these regions are involved in outcome-general as well as outcome-selective PIT (Corbit, 2005; Corbit & Balleine, 2011), the only extant study in humans on appetitive outcome-specific PIT did not find significant involvement of either the nucleus accumbens or amygdala (Bray et al., 2008). Thus, it may be that appetitive PIT BOLD signals in the human amygdala and accumbens are too weak to be observed in paradigms that tap into only outcome-specific or only outcome-general PIT. This could be addressed in future work by increasing the number of trials per subject. Another difference between our and previous work is that we used primary (i.e. appetitive juice) rather than secondary reinforcement (i.e. money) as Pavlovian USs. It is possible that involuntary reception of a juice while lying supine is not as appetitive as receiving money, although our subjective liking ratings did not suggest this was the case. Alternatively, extinction might have been faster for the appetitive than for the aversive juice.

Similar to Talmi et al. (2008), we found that PIT effects were more robust outside than inside the scanner. This replicated attenuation of PIT effects in, but not outside, the scanner, might reflect masking by non-specific factors. The scanner environment is loud and stressful and may well mask subtle behavioural effects that depend on the display of background stimuli. Nevertheless, it should be noted that we did observe significant behavioural PIT over the group as a whole and, moreover, we also observed significant brain-behaviour associations, strengthening our conclusion that the neural effects relate to behavioural PIT.

Our results suggest that outcome-general PIT involves affective regulation of goal-directed behavioural control systems. This generally concurs with the only PIT study in humans, which has shown that outcome-specific PIT can be sensitive to outcome-devaluation (Allman 2010, see however Hol-

land 2004 for different results in rodents). The current study suggests that, at least in humans, this might also hold for outcome-general PIT.

An understanding of how Pavlovian stimuli influence ongoing behaviour may illuminate important aspects of pathological behaviour. For example, one might conceptualize reactive aggression as seen in many mood (Monahan et al., 2001) or personality disorders (Coccaro et al., 2011) as a potentation of aversive PIT. Aspects of proactive aggression, as seen in psychopathy (Cornell et al., 1996) might on the other hand reflect attenuated aversive PIT. This speaks to the notion that psychopathology could arise not only from abnormality within particular behavioural control systems, such as Pavlovian or goal-directed ones, but also from alterations in their interaction (Huys et al., 2012). Further exploration of these hypotheses will require experiments involving patient groups and precise characterization of interactions between the different behavioural control systems involved. As such, the present study represents a stepping stone to future studies to advance our knowledge on affective, Pavlovian influences over instrumental behaviour.

3.5 REFERENCES

ALEXANDER, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual review of neuroscience, 9, 357–381.

ALLMAN, M. J., DeLeon, I. G., Cataldo, M. F., Holland, P. C., & Johnson, A. W. (2010). Learning processes affecting human decision making: An assessment of reinforcer-selective Pavlovian-to-instrumental transfer following reinforcer devaluation. Journal of Experimental Psychology: Animal Behavior Processes, 36(3), 402–408.

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

BIJTTEBIER, P., Beck, I., Claes, L., & Vandereycken, W. (2009). Gray's Reinforcement Sensitivity Theory as a framework for research on personality-psychopathology associations. Clinical Psychology Review, 29(5), 421–430.

BRAINARD, D. H. (1997). The Psychophysics Toolbox. Spatial vision, 10(4), 433–436.

BRAY, S., Rangel, A., Shimojo, S., Balleine, B., & O'Doherty, J. P. (2008). The neural mechanisms underlying the influence of pavlovian cues on human decision making. Journal of Neuroscience, 28(22), 5861–5866.

BÜCHEL, C., Wise, R. J., Mummery, C. J., Poline, J. B., & Friston, K. J. (1996). Nonlinear regression in parametric activation studies. NeuroImage, 4(1), 60–66.

CLARK, L., & Manes, F. (2004). Social and Emotional Decision-making Following Frontal Lobe Injury. Neurocase, 10(5), 398–403.

COCCARO, E. F., Sripada, C. S., Yanowitch, R. N., & Phan, K. L. (2011). Corticolimbic function in impulsive aggressive behavior. Biological Psychiatry, 69(12), 1153–1159.

CORBIT, L. H., & Balleine, B. W. (2005). Double Dissociation of Basolateral and Central Amygdala Lesions on the General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer. Journal of Neuroscience, 25(4), 962–970.

CORBIT, L. H., & Balleine, B. W. (2011). The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. Journal of Neuroscience, 31(33), 11786–11794.

CORNELL, D. G., Warren, J., Hawk, G., Stafford, E., & al et. (1996). Psychopathy in instrumental and reactive violent offenders. Journal of Consulting and Clinical Psychology, 64(4), 783–790.

DAMASIO, A., & Everitt, B. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex [and discussion]. Phil. Trans. R. Soc. Lond..

DAMASIO, A. R. (1997). Towards a neuropathology of emotion and mood. Nature, 386(6627), 769–770.

DAW, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature Neuroscience, 8(12), 1704–1711.

DAYAN, P., & Huys, Q. J. M. (2008). Serotonin, inhibition, and negative mood. PLoS computational biology, 4(2), e4.

DAYAN, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. Neural networks : the official journal of the International Neural Network Society, 19(8), 1153–1160.

DI Giusto, J. A., Di Giusto, E. L., & King, M. G. (1974). Heart rate and muscle tension correlates of conditioned suppression in humans. Journal of Experimental Psychology, 103(3), 515–521.

EVANS, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. Annual review of psychology, 59, 255–278.

FLAGEL, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czuj, A., Willuhn, I., et al. (2011). A selective role for dopamine in stimulus-reward learning. Nature, 469(7328), 53–57.

GITELMAN, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. NeuroImage, 19(1), 200–207.

GREENE, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. Science, 293(5537), 2105–2108.

HABER, S. N. (2003). The primate basal ganglia: parallel and integrative networks. Journal of chemical neuroanatomy, 26(4), 317–330.

HABER, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. Journal of Neuroscience, 20(6), 2369–2382.

HABER, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology, 35(1), 4–26.

HARE, T. A., Camerer, C. F., Knoepfle, D. T., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. Journal of Neuroscience, 30(2), 583–590.

HARE, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. Science, 324(5927), 646–648.

HOLLAND, P. C. (2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. Journal of Experimental Psychology: Animal Behavior Processes, 30(2), 104–117.

HOMAYOUN, H., & Moghaddam, B. (2009). Differential representation of Pavlovian-instrumental transfer by prefrontal cortex subregions and striatum. European Journal of Neuroscience, 29(7), 1461–1476.

HUYS, Q. J. M., Cools, R., Golzer,¨ M., Friedel, E., Heinz, A., Dolan, R. J., et al. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. PLoS computational biology, 7(4), e1002028.

HUYS, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. PLoS computational biology, 8(3), e1002410.

KABLE, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. Neuron, 63(6), 733–745.

KAHNT, T., Heinzle, J., Park, S. Q., & Haynes, J.-D. (2011). Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. NeuroImage, 56(2), 709–715.

LIM, S.-L., O'Doherty, J. P., & Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. Journal of Neuroscience, 31(37), 13214–13223.

MAYBERG, H. S., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R. K., Jerabek, P. A., et al. (1999). Reciprocal Limbic-Cortical Function and Negative Mood: Converging PET Findings in Depression and Normal Sadness. American Journal of Psychiatry, 156(5), 675–682.

MONAHAN, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., et al. (2001). Rethinking risk assessment: The macarthur study of mental disorder and violence. OUP.

ONGÜR, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cerebral cortex (New York, N.Y. : 1991), 10(3), 206–219.

PATENAUDE, B., Smith, S. M., Kennedy, D. N., & Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage, 56(3), 907–922.

POSER, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. Magnetic resonance in medicine, 55(6), 1227–1235.

RANGEL, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. Nature Reviews Neuroscience, 9(7), 545–556.

RUSHWORTH, M. F. S., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal Cortex and Reward-Guided Learning and Decision-Making. Neuron, 70(6), 1054–1069.

SHAROT, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. Nature, 450(7166), 102–105.

TALMI, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human pavlovian-instrumental transfer. Journal of Neuroscience, 28(2), 360–368.

TREW, J. L. (2011). Exploring the roles of approach and avoidance in depression: an integrative model. Clinical Psychology Review, 31(7), 1156–1168.

TVERSKY, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211(4481), 453–458.

TZOURIO-MAZOYER, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage, 15(1), 273–289.

VALENTIN, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. Journal of Neuroscience, 27(15), 4019–4026.

WALLIS, J. D. (2007). Neuronal Mechanisms in Prefrontal Cortex Underlying Adaptive Choice Behavior. Annals of the New York Academy of Sciences, 1121(1), 447–460.

WEINSTEIN, N. (1980). Unrealistic optimism about future life events. Journal of personality and social psychology, 39(5), 806–820.

WIT, S. de, Corlett, P. R., Aitken, M. R., Dickinson, A., & Fletcher, P. C. (2009). Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. Journal of Neuroscience, 29(36), 11330–11338.

WORSLEY, K. J., & Friston, K. J. (1995). Analysis of fMRI Time-Series Revisited— Again. NeuroImage, 2(3), 173–181.

YIN, H. H., Knowlton, B. J., & Balleine, B. W. (2005). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning.

THE European journal of neuroscience, 22(2), 505–512.

YIN, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. The European journal of neuroscience, 22(2), 513–523.

Part III

NEUROCOGNITIVE MECHANISMS UNDERLYING AFFECTIVE DYSREGULATION OF BEHAVIOUR IN PATIENTS

4

AMYGDALA RESPONSE PREDICTS CLINICAL SYMPTOM REDUCTION IN PATIENTS WITH BORDERLINE PERSONALITY DISORDER

Geurts, D.E.M., Van den Heuvel, T.H., Verkes R.J., Huys Q., Cools R.

Submitted.

Background: Borderline personality disorder (BPD) is a prevalent, devastating and heterogeneous psychiatric disorder. Psychotherapeutic treatment success is highly variable within this patient group. Neurobiological mechanisms might mitigate phenomenological heterogeneity while also providing us with predictors of treatment success. Here we build on previous observations that BPD is accompanied by enhanced impact of aversive affect on behavior and abnormal neural signaling in the amygdala. Aim: To assess whether BPD is accompanied by abnormal aversive regulation of instrumental behavior and associated neural signaling, in a manner that is predictive of symptom reduction after therapy. **Method:** We tested a clinical sample of 15 patients with BPD, awaiting dialectical behavioural therapy, and 16 matched healthy controls using an aversive Pavlovian-to-instrumental transfer (PIT) task that assesses how instrumental approach and withdrawal behaviors are influenced by aversive Pavlovian conditioned stimuli (CSs). All participants were scanned with fMRI during the critical Pavlovian-to-instrumental transfer phase. Patients were assessed 1 year after the start of dialectical behavioural therapy to quantify changes in BPD symptom severity. **Results:** At baseline, aversive PIT and associated neural signaling did not differ between groups. However, BOLD signal in the amygdala measured during aversive PIT predicted symptom reduction at 1 year follow-up: Enhanced aversive amygdala signaling before treatment was highly associated with reduced clinical improvement after 1 year of treatment. **Conclusion:** Clinical symptom reduction over 1 year of treatment in BPD patients can be predicted from BOLD signal in the amygdala, measured using an aversive PIT task. This finding demonstrates a key role for the amygdala in the recovery of borderline personality disorder. The results suggest that excessive responsiveness of the amygdala during aversive PIT might render patients resistant to symptom improvement.

4.1 INTRODUCTION

Borderline personality disorder (BPD) is a prevalent and devastating psychiatric disorder. BPD patients struggle with pervasive, persistent and profound interpersonal, affective and impulse regulation problems. The disorder is associated with high levels of treatment utilization, severe functional impairments and high mortality rates (American psychiatric association, 2000; Grant et al., 2008; Bolton and Robinson, 2010). Lifetime prevalence is as high as 6% in a large community sample (Grant et al., 2008) and a staggering 20% among psychiatric inpatients (American psychiatric association, 2000). Apart from the suffering of BPD patients and their families, costs for society are high due to heavy use of expensive health care resources and the persistent lack of productivity (Wunsch et al., 2014). Optimizing care for this patient group is hence of major importance (Gunderson, 2009).

Although several psychotherapeutic treatments exist for BPD, response is highly variable and treatment effects are modest overall (Stoffers et al., 2012). For example 27-35% of patients continue to have admissions, self-harm and conduct suicidal gestures (Lana and Fernández-San Martín, 2013). Neurocognitive mechanistic research might help us identify key predictors of treatment success and thus mitigate the large variability in treatment efficacy (Jones et al., 2015; Heinz et al., 2016). However, remarkably little work has been done to investigate potential associations between clinical symptom reduction and neurocognitive functioning prior to treatment in patients with BPD (though see Perez et al., 2015). We fill this gap by assessing symptom reduction over one year in terms of pre to post change in borderline personality severity as a function of neurocognitive processing, measured prior to the start of 1 year of dialectical behavior therapy (DBT).

We focused on affective dysregulation of instrumental behavior. Despite being a central affective/cognitive aspect of BPD it has not received its deserved empirical attention. Aversive affective states have long been recognized to have an excessive negative impact on behavior in BPD patients (Linehan, 1993; Selby et al., 2009). Indeed, the interaction between aversive affect and anomalous behavior is one of the main foci of DBT, a comprehensive psychological therapy for BPD. This therapy focuses on the excessive impact of especially negative emotions on behavior and how these can be accepted and dealt with skillfully through mindfulness, skills training, distress tolerance and emotion regulation. In other words, the core aim of this therapy is to optimize the interaction between affect and rational behaviour. Despite this focus of the therapy on the impact of aversive affect on behavior, experimental work has addressed primarily affective processing per se,

rather than the consequences of abnormal affective processing for behavior (but see for example Silbersweig et al., 2007; Jacob et al., 2013; for review see Rosenthal et al., 2008). In the present study we used a previously validated laboratory task during fMRI to assess whether DBT alters the affective regulation of instrumental behavior, and associated neural signaling, in BPD patients.

Specifically, we targeted the interaction between aversive affect and behavior by measuring aversive Pavlovian to instrumental transfer (PIT). Aversive PIT refers to the observation that aversive instrumental actions, such as inhibition and withdrawal, are potentiated in the context of aversive Pavlovian cues, i.e. stimuli that predict aversive outcomes. Thus aversive Pavlovian cues have been shown to inhibit instrumental approach actions (this form of aversive PIT is known as conditioned suppression) and to enhance instrumental withdrawal actions (Huys et al., 2011; Geurts et al., 2013a). Accumulating evidence from studies with experimental animals and more recently also healthy humans (Talmi et al., 2008; Allman et al., 2010; Huys et al., 2011; Prevost et al., 2012; Geurts et al., 2013b; Lewis et al., 2013; Geurts et al., 2013a; Hebart and Gläscher, 2015; Watson et al., 2014) and patients (Garbusow et al., 2015) demonstrates involvement of (fronto)limbic circuitry in PIT, including the amygdala (Cardinal et al., 2002; Talmi et al., 2008; Balleine and Doherty, 2009; Prevost et al., 2012; Geurts et al., 2013a; Ly et al., 2014). This is particularly relevant in the context of the current study, because the amygdala has also been central to neurocognitive theories and empirical research on BPD. For example a recent meta-analysis reports functional hyperactivity of the left amygdala during aversive versus neutral stimuli as well as smaller gray matter volume of the amygdala in BPD (Schulze et al., 2016). This amygdala hyperactivation has been proposed to reflect deviant salience of negative emotional stimuli and to be remediated by psychotropic medication in BPD (Schulze et al., 2016). Furthermore preliminary evidence shows that effects of DBT are also associated with changes in blood oxygen level dependent (BOLD) signal in this region (Schnell and Herpertz, 2007; and later: Goodman et al., 2014). Here we build on these previous findings by assessing the hypothesis that BPD is accompanied by abnormalities in aversive PIT and associated BOLD signal in the amygdala. Moreover, we ask whether aversive PIT and associated amygdala signal might be predictive of symptom reduction after DBT treatment.

A previously established PIT paradigm was employed that assesses how instrumental approach and withdrawal behaviors are influenced by aversive Pavlovian conditioned stimuli (CSs)(Geurts et al., 2013a). Although extant data suggest that the amygdala is central to PIT, these underpinnings

are not confined to the (peri)amygdalar region. Previously, we found that BOLD responses in the amygdala and nucleus accumbens were associated with behavioral inhibition by aversive Pavlovian cues. Moreover we found that this inhibition was associated with the extent to which aversive cues modulated connectivity between the ventromedial prefrontal cortex and the (dorsomedial and ventral) striatum. This led us to extend our analyses beyond the amygdala to the ventromedial prefrontal cortex and the striatum. Indeed, evidence from functional, structural and metabolic studies with BPD patients indicates involvement of an extended fronto-limbic brain network (Krause-Utz et al., 2014; e.g. Salvador et al., 2016).

First we investigated baseline task performance differences between healthy controls and BPD patients. Based on the evidence reviewed above (Linehan, 1993; Selby et al., 2009; Schulze et al., 2016), we hypothesized that, relative to controls, BPD patients would exhibit excessive impact of aversive Pavlovian CSs on instrumental behavior as well as enhanced BOLD signaling in the amygdala and fronto-striatal circuitry. The study also allowed us to assess the alternative hypothesis that BPD is accompanied by reduced rather than enhanced aversive PIT. This alternative hypothesis would concur generally with recent findings that stress, a key characteristic of BPD, might reduce rather than enhance PIT (Quail et al., 2016, but see Pool, 2015). Critically, we also predicted that aversive PIT-related BOLD-signal in the BPD group would predict symptom reduction after one year of DBT treatment.

4.2 MATERIALS AND METHODS

4.2.1 *Participants*

To maximize external validity we aimed for a patient sample that would represent patients treated in general mental health practice as closely as possible (Hoertel et al., 2015). Therefore all patients who were enrolled in the pre-treatment phase of an already existing one year DBT program at the Radboud University Medical Centre between March 2012 and March 2013 (n=29) were asked to participate in this study. Twenty-three patients volunteered. Imaging datasets were obtained for 15 patients and clinical outcome measures after treatment were obtained for 14 of these patients (see supplement 4.7.1). In addition, 16 healthy controls matched for gender, age and education were recruited per advertisement (for group demographics and questionnaire scores see Table 4.1, for comorbidity and medication use of the BPD group, see Supplementary table 4.2).

4.2.2 *Procedure*

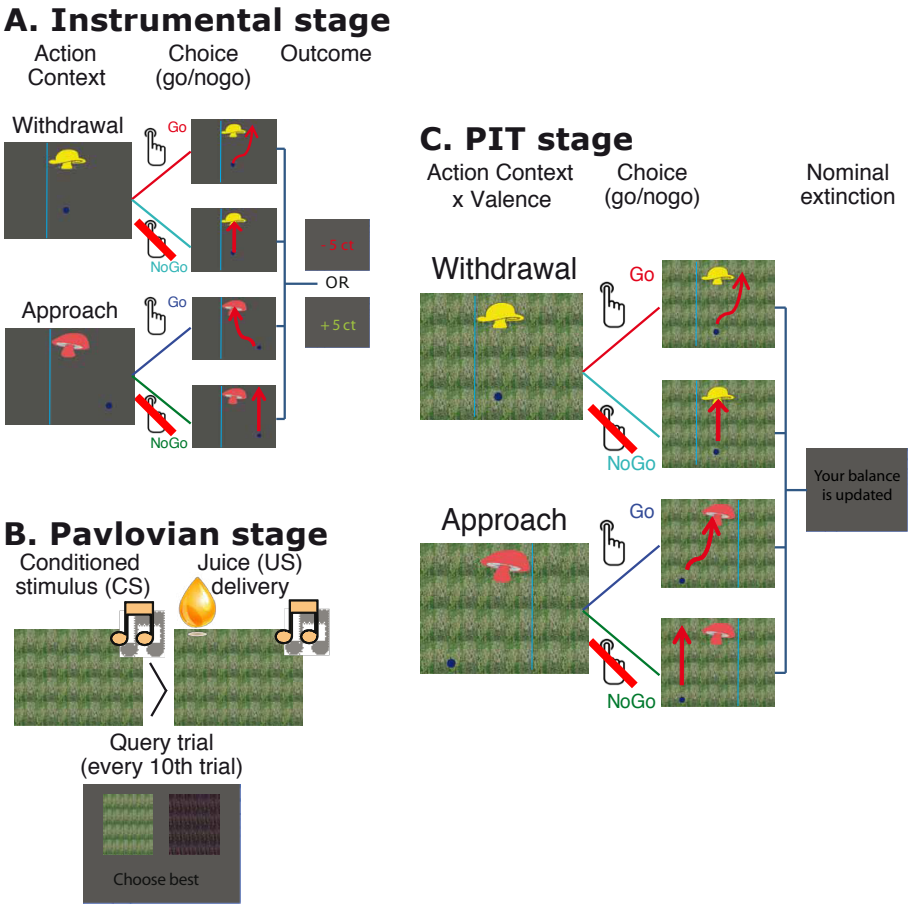
All patients enrolled in the pre-treatment phase of the DBT were invited to attend 3 sessions: First, a screening session; second, a pre-treatment scan session; and third, a post-treatment assessment. During the screening session, participants received a full diagnostic structured interview, which included the MINI-plus international neuropsychiatric interview and the Structured Clinical Interview for DSM-IV Axis II disorders (SCID-II), administered by a senior resident in psychiatry (author DG). The local Medical Ethical Committee approved the study (NL36001.091.11). To familiarize subjects during the first visit with the scanning environment and procedures we employed a short scan-session of about 15 minutes during which a structural MRI scan was obtained and subjects were familiarized with the instructions and instrumental and Pavlovian training stages in the scanner. During the second visit, just before treatment started, subjects completed several questionnaires (Table 4.1) of which the Borderline Personality Disorder Checklist (BPD47) measuring the disease severity was of primary interest. Before entering the scanner, instructions on the computer task were repeated orally. After receiving the instructions a third time, now projected on the scanner screen, they started the PIT paradigm. After a 15 minute break subjects performed a short neuropsychological test battery (Table 4.1). The third and final follow-up session followed approximately one year later and no earlier than treatment had ended. Subjects completed the same questionnaires and participated in the same neuropsychological test battery as in the second session (Table 4.1). In addition the MINI was administered once again to investigate whether axis I classifications had changed and the BPD47 to measure changes in borderline symptom severity.
]

4.2.3 *Pavlovian-instrumental transfer paradigm*

Participants performed a computerized task to assess aversive PIT (Geurts et al., 2013a). The experiment consisted of three stages: (1) instrumental conditioning, (2) Pavlovian conditioning and (3) PIT. The instrumental stage contained two Action contexts: (i) a context in which the active response led to an approach action and (ii) another in which the active response led to a withdrawal action. In each context 2 go-stimuli and 2 nogo-stimuli were repeatedly presented to the participant (Figure 4.1A). In the approach Action context participants learned through monetary feedback (wins and losses) whether to 'collect' the instrumental stimulus (approach-go) or not

	Healthy controls		Borderline personality disorder group			
	Baseline				1 year follow-up	
	N= 16		N = 15		N=14	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Age	29.5	8.8	28.5	8.8	-	-
IQ (NLV)	101.8	12.3	100.3	11.5	-	-
BPD47	6.7	6.5	79.7	33.2	64.8	30.0
OQ – total	42.5	20.6	91.5	19.2	79.1	22.5
Sympt. distr.	19.1	10.6	56.7	14.1	50.0	16.5
Inter. pers.	8.8	4.9	20.2	3.8	17.5	5.4
Social role	8.8	4.2	15.0	4.5	11.6	3.6
BDI-II	3.6	4.0	33.4	14.3	28.4	14.0
BIS	18.4	7.1	23.5	4.1	24.4	3.8
BAS	38.7	14.7	40.1	5.8	41.4	5.1
Box Completion (s)	85.4	30.7	107.0	20.9	96.8	27.4
Digit Span	13.2	2.5	16.2	4.0	15.2	4.1
Forward	7.1	1.6	8.3	1.9	7.7	2.2
Backward	6.0	1.3	7.9	2.4	7.5	2.4
Verbal Fluency	44.6	12.1	38.1	11.4	42.3	10.3

Table 4.1 *– Demographical and clinical characteristics of the borderline personality disorder and healthy matched control participants. SD, standard deviation; NLV, Dutch reading test; BPD47, Borderline personality disorder checklist; OQ, outcome questionnaire; BDI-II, Beck depression index 2nd version; BIS, behavioural inhibition systems; BAS, behavioral activation system.*



(approach-no-go). In the withdrawal Action Context they learned to avoid collecting instrumental stimuli (withdrawal-go) or not (withdrawal-no-go). Instrumental stimuli were randomly assigned to 1 of the 4 trial types. Thus, in both the approach and withdrawal Action Contexts, there were 2 go-stimuli, which yielded reward more often (i.e. 80% of the cases) after active responses (and punishment after not responding), and 2 nogo-stimuli, which yielded reward more often (i.e. also 80% of the cases) after not responding (and punishment after go-responding).

The second, Pavlovian stage consisted of repeated presentation of three audiovisual stimuli (Figure 4.1B): The appetitive and aversive con-

Figure 4.1 *– Task details.* ***A. Instrumental stage.*** *Each instrumental stimulus was presented 10 times. Trials started with the appearance of the instrumental stimulus at the top center of the screen and a dot at the bottom of the screen. In approach trials, the dot appeared either on the left or on the right bottom of the screen. Participants could choose to do nothing (approach-no-go), in which case the dot would move past the instrumental stimulus. Alternatively, they could press the button repeatedly to steer the dot through the instrumental stimulus (approach-go). In withdrawal trials, the dot started centrally beneath the instrumental stimulus. Participants could choose to press the button repeatedly to avoid moving through instrumental stimulus (withdrawal-go) or to do nothing (withdrawal-no-go). The four possible trajectories are drawn in the figure (red lines). If the dot entered the target region, then the instrumental stimulus was 'collected'. The vertical line to one side of the instrumental stimulus could not be crossed by the dot. Pressing the button led always to approach in the approach context, and to withdrawal in the withdrawal context. Timings were as follows: Instrumental stimuli were presented for 2.5 sec, during which responses were collected. After this, feedback was presented for 1 sec. The intertrial interval (ITI) was 1 sec (blank screen). There were 80 instrumental trials divided in miniblocks of 8 withdrawal or approach trials.* ***B. Pavlovian stage*** *. Each Pavlovian CS was presented 20 times. Stimulus presentation order was fully randomized across participants. Stimulus duration was 4.5 sec, and juice delivery (2ml) occurred between 0 and 1.5 sec after stimulus onset. The ITI was 1 sec. Query trials were presented after every 10 Pavlovian trials. On these trials, participants were instructed chose the best of the two presented Pavlovian stimuli (presented for 2 sec; ITI 0.5 sec) without any feedback.* ***C. PIT stage.*** *The PIT stage paralleled the instrumental training, except that Pavlovian CSs tiled the background. Each instrumental stimulus was presented 12 times and each Pavlovian CS 32 times counterbalanced over the 8 instrumental stimuli. No outcomes were presented in this phase, but participants were instructed that their choices counted toward the final total (known as nominal extinction). Participants were explicitly instructed that the juices were collected outside the scanner, and they agreed before the start of the experiment to drink them afterward. Timing of a single trial was as follows: 0.25 sec after the onset of the Pavlovian stimulus, the instrumental stimulus (and dot) was overlaid on top of this Pavlovian stimulus. Duration of the instrumental stimulus was 2.5 sec; duration of the Pavlovian stimulus was 2.75 sec. Upon offset of both stimuli, feedback was presented, which consisted only of the words "Your balance is updated" (duration = 1 sec, ITI = 1 sec).*

ditioned stimuli (CS) were followed, respectively, by appetitive or aversive juice (i.e. the unconditioned stimuli USs) on 50% of trials. The neutral CS resulted in no outcome. The appetitive juice was based on subjective preference for apple, orange or strawberry lemonade. The aversive juice was a bitter magnesium sulphate solution (0.3M). Conditioning was assessed in two ways: (1) participants indicated the degree to which they liked each of the CSs (and USs) by use of visual analogue scales (VAS), before and after the experiment; (2) participants chose one of the two presented Pavlovian stimuli (presented for 2s; ITI 0.5s) in extinction on 12 interspersed query trials.

In the third, PIT, stage stimulus presentation was the same as in the instrumental stage, except that Pavlovian stimuli tiled the background from 250ms before (Larson et al., 2013) and during the instrumental trial, and no outcomes were presented (Figure 4.1C). Participants were instructed that their choices counted towards the final monetary total, and that the juices associated with the Pavlovian outcomes were collected outside the scanner for them to drink afterwards. There were 2 independent runs separated by a 2 minute break (each including run-specific stimuli/CSs), with each run including all three stages.

4.2.4 *Image acquisition*

Whole-brain imaging was performed on a 1.5 Tesla MR scanner (Avanto, Siemens Medical Systems, Erlangen, Germany). Functional data were obtained using a multi-echo gradient T2*-weighted echo-planar (ME-EPI) scanning sequence (Poser et al., 2006)(Supplementary Materials).

4.2.5 *Analysis*

Our primary analysis was restricted to the PIT stage. Results on the instrumental and Pavlovian training data are presented in the supplement (4.7.2). The analyses presented below consist of two parts: First, we assessed effects of group on behavior and fMRI BOLD response during the PIT stage, measured at baseline. Second, within the BPD group we assessed whether aversive PIT and associated BOLD signal were associated with symptom reduction after 1 year of treatment.

Pavlovian-instrumental transfer

BEHAVIORAL ANALYSES Effects of Action Context (approach/withdrawal), CS Valence (neutral/aversive) and Group (healthy controls/BPD patients) in the critical transfer test were assessed in terms of proportion of go-choices (p(go), not normally distributed) and average number of button presses made during these go-choices. Analyses were targeted at aversive PIT effects, that is the degree to which aversive CS inhibited instrumental 'go' responding (contrast [p(go | neutral)- p(go | aversive)]). In addition, we also assessed the Action Context-specificity of aversive PIT, with the contrast [p(go | neutral,approach) - p(go | aversive,approach) – (p(go | neutral,withdrawal) - p(go | aversive,withdrawal))]. Note that we focused our analyses on aversive PIT, based on our hypothesis (see introduction) and on our previous work (n=33) showing that the current paradigm was not sensitive to (and therefore not valid to assess group effects on) appetitive PIT (Geurts et al., 2013a).

FMRI ANALYSIS fMRI analysis was performed with SPM5 software (Wellcome Trust Centre for Cognitive Neuroimaging, London, UK). Preprocessing steps and first-level fMRI analysis were identical to (Geurts et al., 2013a): We first applied a PAID-weight algorithm(Poser et al., 2006) to optimally combine the different echoes, slice-time correction, coregistration, normalization based on parameters estimated through segmentation of the structural images, and smoothing (8 mm FWHM). The fMRI analysis was restricted to the PIT-stage, and was similar to our previous analyses (Geurts et al., 2013a). The general linear model (GLM) at the participant level consisted of six main regressors (four of interest) representing the onset of the six different PIT trials (Action Context (approach/withdrawal) x CS Valence (appetitive/neutral/aversive)). For each main regressor an additional parametric regressor was added(Büchel et al., 1996): The PIT-regressor (Talmi et al., 2008; Geurts et al., 2013a) was a parametric modulator of BOLD responses by the number of button presses per trial. Contrasting this regressor between the different CS-valence thus reveals regions where BOLD signal is associated with valence-dependent coupling between amygdala BOLD signal and instrumental behaviour on a trial by trial basis. Note that such a contrast goes beyond simple reactivity of region to a CS or to instrumental behavior; it critically captures its interaction, i.e. PIT.

A further parametric regressor contained the expectation associated with each instrumental stimulus (the Q-value) per trial as estimated from a model-based analysis of behavior (Huys et al., 2011). This was done based on prior data showing that BOLD signal in the prefrontal cortex and

striatum, our regions of interest, covaries with instrumental action value (Valentin et al., 2007; Wunderlich et al., 2009). As such, this approach maximized the degree to which our GLM captured variability in relevant BOLD signal. Furthermore, realignment parameters were added, high-pass filtering (128s) was applied and parameter estimates were obtained by maximum-likelihood estimation (AR1).

The parameter estimates for the neutral and aversive parametric PIT-regressors were used in a 2x2x2 rmANOVA at the group-level (with random effects) with Action Context (approach/withdrawal) and Valence (neutral/aversive) as within-participant factors and Group (healthy controls/BPD) as between-participants factor. Within this rmANOVA we assessed the main effect of Action Context and based on Geurts et al. (2013a) we expected this analysis to reveal that BOLD signal in the ventromedial prefrontal cortex would be Action Context specific (approach > withdrawal). We did not expect a group effect on this contrast.

To capture additional PIT signal that is related to stable patterns of behavior beyond trial-by-trial variation in instrumental vigour we contrasted the main regressors(Talmi et al., 2008; Geurts et al., 2013a) at the participant-level to calculate the main effect of Valence [(approach neutral+withdrawal neutral) - (approach aversive+withdrawal aversive)]. The resulting SPM was then used in a two-sample t-test at the group-level with aversive PIT in terms of the average number of button presses as a covariate for each group separately enabling comparison between groups (Geurts et al., 2013a). Based on Geurts et al. (2013a) we expected that behavioral aversive PIT in terms of the average number of button presses [(BP | approach&neutral+BP | withdrawal&neutral) - (BP | approach&aversive+BP | withdrawal&aversive)] would be related to BOLD signal change (neutral-aversive) in the amygdala and nucleus accumbens. Here, we anticipated group differences in neural underpinnings of PIT especially in the amygdala (see 4.2 introduction). We planned to use significant clusters revealed by the previous analyses (if any) as seeds in a generalized psychophysiological interaction analyses(McLaren et al., 2012) to assess task related functional connectivity with other brain regions (supplementary analysis).

Treatment success and its prediction

Our primary measure of treatment success was the Borderline Personality Disorder Checklist (BPD47: Giesen-Bloo et al., n.d.) a 47 item self-report questionnaire based on the Borderline Personality Disorder Severity Index by Arntz et al. (2003). Furthermore, as secondary measures we also assessed

quality of life with the Outcome Questionnaire and depressive symptoms with the Beck Depression Inventory second edition (BDI-II). Treatment success was computed by subtracting the post-treatment scores from those acquired during the first scan session.

Predictive relationship between aversive PIT and symptom reduction We assessed the association between aversive PIT and associated BOLD (at the whole-brain level and within the predefined amygdala ROI), measured pre-treatment, with clinical symptom reduction 1 year later. A second-level random effects simple regression analysis was conducted on Action Context specific and aversive PIT-related BOLD signal ([PIT regressor | approach&neutral - PIT regressor | withdrawal&neutral) - (PIT regressor | approach&aversive - PIT regressor | withdrawal&aversive)] and [PIT regressor | neutral - PIT regressor | aversive)] respectively) with BPD47 change (before-after) as the covariate of interest. In addition, we also performed the non-parametric equivalent of this analysis with SnPM and we employed a leave one participant out procedure (Esterman et al., 2010), in which a single participant is iteratively left out of the second level correlational analysis (i.e. simple regression analysis of the aversive PIT statistical parametric maps with the covariate of symptom reduction). The resulting clusters within the anatomically defined bilateral amygdala (thresholded at p<.001 uncorrected) were then used to extract the mean beta weights of the left out participant to calculate the aversive PIT contrast. This procedure was repeated for each participant. The GLM from the remaining participants thus serves as an independent localizer for the participant left out (Esterman et al., 2010).

Statistical thresholding

We report effects that survive family wise error (FWE) correction for multiple comparisons across the whole brain (P_{WB} <.05, voxel-level) or in one of the following regions of interest (ROIs): The amygdala (automated anatomical labeling atlas, Tzourio-Mazoyer et al., (2002)) was our primary ROI to assess the prediction of symptom reduction. Both the amygdala and nucleus accumbens were chosen as ROIs for the analysis of the main PIT task effects (across and between groups) based on their key role in PIT (Corbit, 2005; Talmi et al., 2008; Corbit and Balleine, 2011; Prevost et al., 2012; Geurts et al., 2013a). Specifically, in our previous study we found BOLD response in both these regions to be associated with behavioral PIT on a participant by participant basis. The caudate nucleus was chosen based on its general role in controlling both motivation and instrumental action (Cardinal et al., 2002) and its specific role in connection with the vmPFC in PIT (Geurts et

al., 2013a). Finally, following our prior work, we also assessed Action Context specificity in the ventromedial prefrontal cortex: The region shown to be sensitive to Action Context in our previous PIT study was used as ROI (Geurts et al., 2013a). The left and right elements of each bilateral volume of interest were combined using Marsbar™(Brett et al., 2002).

4.3 RESULTS

4.3.1 *Baseline behavioral data*

Pavlovian-instrumental transfer

The main task effects were as previous. Thus, consistent with our previous studies using this paradigm (Huys et al., 2011; Geurts et al., 2013a), we observed opposite effects of the aversive Pavlovian CSs on approach and withdrawal actions (in terms of choice p(go), Figure 4.2): Planned contrasts confirmed the statistical significance of this action specificity of the aversive PIT effect (Related-Samples Wilcoxon Signed rank test [p(go | approach,neutral) - p(go | approach,aversive)] > [p(go | withdrawal,neutral)- p(go | withdrawal,aversive)]: p=.031, one-tailed). There were no differences between the groups (Independent samples median test: p=.48) but we note that the action-specific PIT effect was present in healthy controls (p=.008), but not in patients (p=.860) when examined separately.

There were no main task effects except for a main effect of Action Context (($F_{(1,29)}$=33.7, p<.001, all other F<1.8 and p>0.2) in terms of vigour (average number of button presses; Supplementary table 4.3). There were also no group differences.

Performance on the instrumental task and assessments of Pavlovian training did not differ between the groups (Supplementary results).

4.3.2 *Baseline imaging data*

Consistent with our previous fMRI study using this paradigm (Geurts et al., 2013a), BOLD signal in the vmPFC varied as a function of Action Context. Signal was greater during approach than during withdrawal (small volume corrected results for the vmPFC ROI: peak voxel MNI-coordinates [-6 32 -12], k=45, Z=3.86, p_{FWE}=.021, Figure 4.3). There were no group differences as a function or independent of Action Context. Conversely, we did not replicate the previously observed correlation between individual

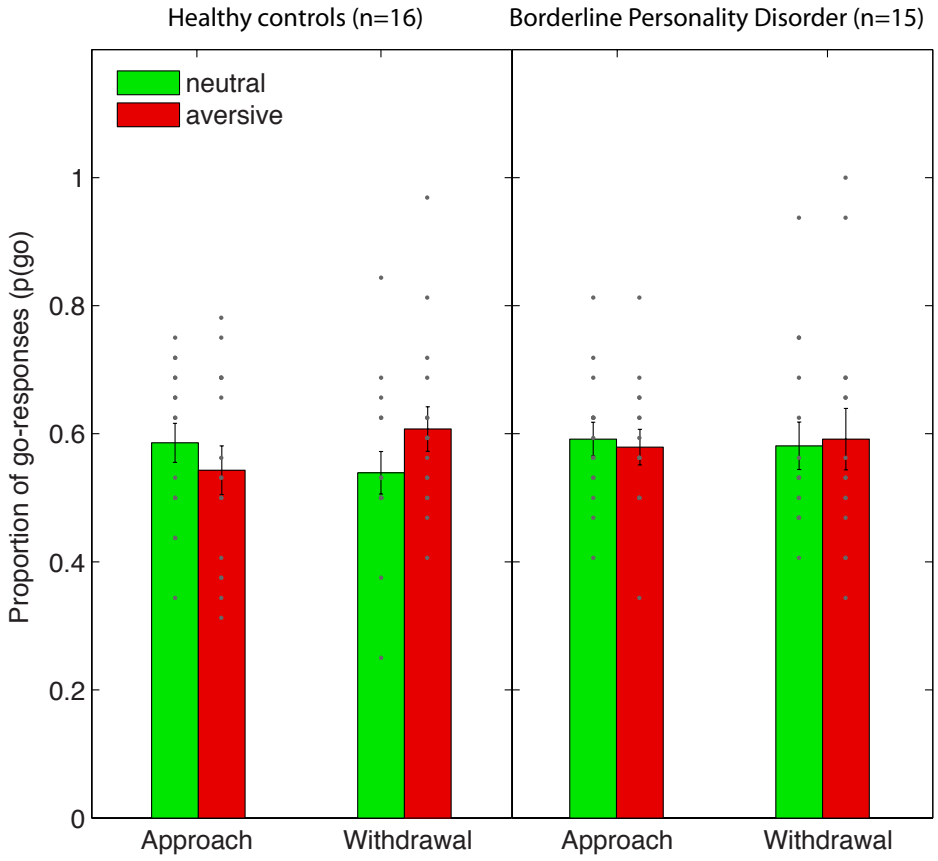


Figure 4.2 *– Behavioral data from the Pavlovian-instrumental transfer stage. Shown are mean proportions of go-responses (p(go)) as a function of Action Context (approach versus withdrawal) and Valence (neutral/aversive). Error bars represent standard errors of the means and dots represent individual data points. Note that there were no significant differences between Groups.*

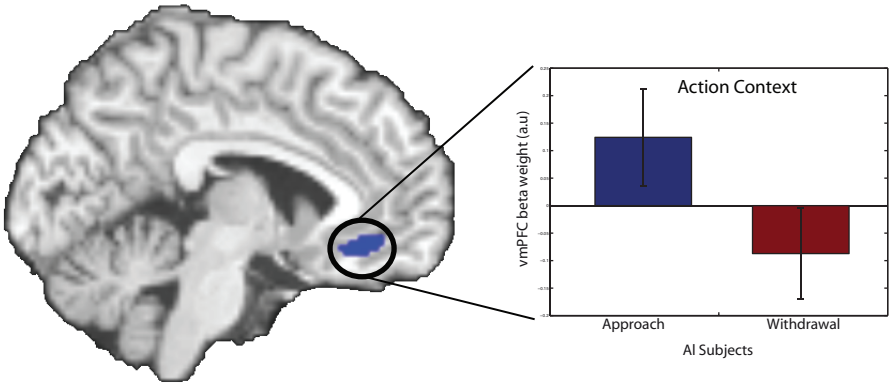


Figure 4.3 *– Action-specific BOLD response in the vmPFC. There was a main effect of Action Context in the vmPFC (peak voxel MNI-coordinates [-6 32 -12], k=45, Z=3.86, p_{FWE}=.021, small volume corrected). The bar graph shows parameter estimates from the peak voxel for the different Action Contexts. Images are displayed at a statistical threshold of p < .001 uncorrected*

differences in behavioral aversive PIT and BOLD signal in the amygdala and nucleus accumbens (Geurts et al., 2013a). There was also no evidence for CS dependent functional connectivity between the vmPFC and caudate nucleus (compare with Geurts et al., 2013a). Further analyses relating to this issue are presented in the Supplementary Materials.

4.3.3 *Aversive PIT and symptom reduction*

Symptom reduction

The 14 patients who were seen at follow-up 1 year after the start of therapy showed significant reduction in symptom severity as measured with the BPD47 (mean difference=-17.4, $t_{(13)}$=2.5, p=.027), OQ (mean difference=-12.4, $t_{(13)}$=3.1, p=.009), and in trend with the BDI-II (mean difference=-4.8, $t_{(13)}$=1.8, p=.090).

None of the neuropsychological tests reported in Table 4.1 changed significantly from baseline to 1 year after treatment (all -1.9 > $t_{(13)}$ < 2.2, all p>.05).

PIT-related BOLD signal in the amygdala is related to symptom reduction 1 year later

Pre-treatment BOLD signal in bilateral amygdala was related to BPD symptom reduction after one year (Figure 4.4). Higher aversive PIT-related signals (across Action Context: [approach,aversive] – [approach,neutral] + [withdrawal,aversive] - [withdrawal,neutral]) were inversely associated with symptom reduction 1 year later. This observation was confirmed using both parametric and nonparametric statistical analyses (small volume corrected effects in anatomically defined bilateral amygdala; parametric tests with SPM: peak voxel MNI-coordinates [-24 0 -16], k=22, Z=3.79, p_{FWE}=.027; non parametric test with SnPM: peak voxel MNI-coordinates [-24 0 -18], pseudo-t=4.22, p_{FWE}=.013; and MNI-coordinates [22 4 -18], pseudo t = 3.19, p_{FWE}=.06). The robustness of these effects was confirmed by cross validation ($r_{(14)}$= -.655, p=.011) and by supplementary analyses on mean beta estimates extracted from the anatomically (rather than functionally) defined bilateral amygdala (Pearson $r_{(14)}$= -.667, p=.009).

Next, we examined the specificity of the effect. A stepwise multiple regression analysis revealed that pre-treatment PIT-related signal in the bilateral amygdala ($F_{(1,12)}$ = 9.6, p =.009, multiple correlation coefficient = .67) accounted for variance in symptom reduction over and above the other collected baseline measures (all $|t| < 2.0$, all $p > .05$). Next, we examined whether the predictive effect of pre-treatment PIT-related amygdala signal was specific for BPD47 change or whether it extended to other changes in clinical or neuropsychological measures. Indeed, stepwise multiple regression analysis with this amygdala signal as dependent variable revealed that this signal's association with BPD47 improvement ($F_{(1,12)}$=9.6, p=.009) did not extend to any of the other changes in clinical or neuropsychological measures (all $|t|<2.0$, all $p>.05$). This was relevant, because, unsurprisingly, improvement in borderline severity was accompanied by improvement in depressive symptoms as measured with the BDI-II ($r_{(14)}$=-.67, p=.008). Improvement in BPD47 was also accompanied by improvement in verbal fluency ($r_{(14)}$=.91, p=.000008), and by a change in behavioural activation as measured with the BAS ($r_{(14)}$=-.63, p=.016).

4.4 DISCUSSION

The present study demonstrates that BOLD signal elicited by the aversive PIT effect in the amygdala predicts symptom reduction in patients with borderline personality disorder. Greater PIT-related responsiveness of the

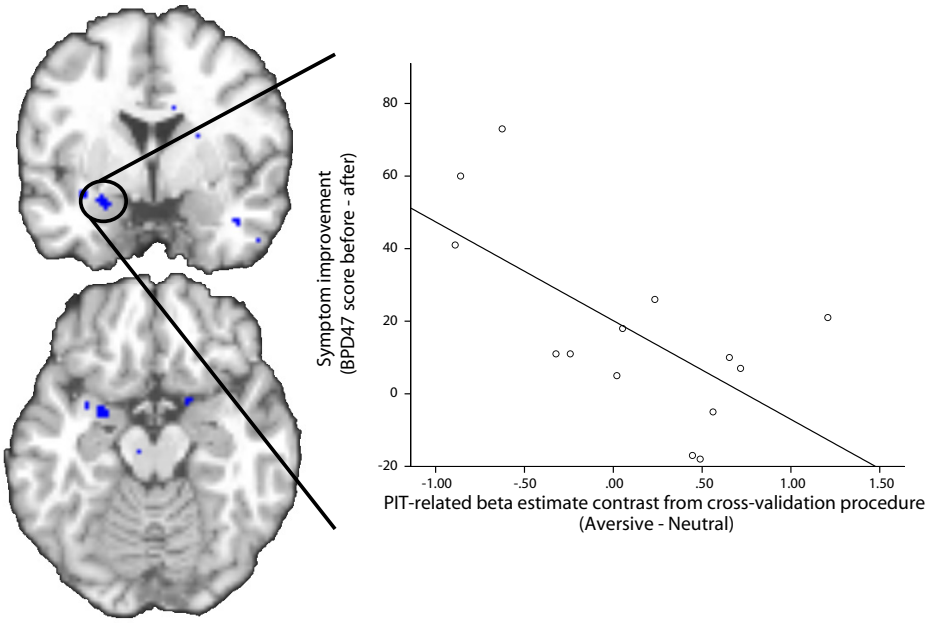


Figure 4.4 *– Association between amygdala BOLD signal change and symptom improvement. Pre-treatment PIT-related BOLD signal in the left amygdala predicts symptom improvement 1 year later. Images are displayed at a statistical threshold of $p < .001$ uncorrected. The scatter plot shows the PIT-related beta estimate contrast for aversive minus neutral CS trials before treatment derived from a leave one participant out cross-validation procedure in relation to symptom improvement. The regression line is the ordinary least square line.*

amygdala was associated with reduced clinical improvement 1 year later. This suggests that individual differences in the degree to which the amygdala responds during aversive Pavlovian to instrumental transfer predict resistance to clinical improvement (or slower recovery) of BPD. Thus, participants who showed increased coupling between amygdala BOLD signal and the suppression of instrumental behavior during aversive trials, showed less clinical improvement.

Based on observations that BPD is associated with enhanced impact of aversive stimuli on behavior (cf. Soloff et al., 2015), we employed an aversive PIT task that measures the degree to which aversive Pavlovian stimuli alter instrumental behavior. We replicated the previously observed basic task effects, including the Action Context-specificity of aversive PIT (Huys et al., 2011; Geurts et al., 2013a), with aversive Pavlovian CSs suppressing approach, but potentiating withdrawal actions. These task effects were not modulated by BPD, although when analyzing the groups separately we only found significant effects in the healthy controls. The absence of a group effect might be due to insufficient power or the relatively stressful scanner environment (Talmi et al., 2008; Geurts et al., 2013a). Indeed, a recent study shows that stress reduces behavioral PIT effects (Quail et al., 2016; but see Pool et al., 2015) and patients with BPD might be more sensitive to this stress. However, we cannot exclude that, as a group, BPD patients indeed do not exhibit abnormal aversive PIT. For example, the absence of an effect of BPD on aversive PIT might well reflect neurocognitive heterogeneity, with abnormal aversive PIT (and associated neural signaling) being present only in a subset of patients.

The importance of such heterogeneity is highlighted by the key observation of this study, showing that individual differences in the amygdala response during aversive PIT predict symptom reduction. Although this provides converging evidence for the validity of this PIT paradigm for predicting clinical symptom changes (in depression: Huys et al., 2016; and addiction: Garbusow et al., 2015), it should be noted that, here, amygdala signal across Action Contexts was the predictor, whereas in the study of Huys et al. (2016) it was the Action Context specificity of behavior that predicted recovery from depression and in the study of Garbusow et al. (2015) it was the PIT-effect in the nucleus accumbens that predicted relapse in alcohol use.. This suggests that different aspects of the neurocognitive mechanisms underpinning the transfer between affective motivation and instrumental behavior might be disorder and/or treatment specific.

The present results suggest that symptom reduction after DBT is greater in BPD patients who show reduced responsiveness of the amygdala during

aversive PIT. The finding that amygdala signal is predictive of symptom reduction in BPD after DBT concurs with empirical findings and neurocognitive theories implicating a central role for the amygdala in BPD (Schulze et al., 2016) and DBT (Schnell and Herpertz, 2007; Goodman et al., 2014). Two recent studies assessed changes in neurocognitive processing during DBT (Schnell and Herpertz, 2007; Goodman et al., 2014). Schnell et al. employed a pilot study with 6 BPD patients who received several fMRI scans during 3 month DBT. The 4 patients who responded to DBT all showed decreases in amygdala BOLD responses to emotional pictures. In keeping with this finding, Goodman et al. (2014) reported decreases in amygdala responses to emotional picture and associated improvement in self-reported emotional regulation in eleven BPD patients after 1 year of DBT treatment. These studies suggest that the association between amygdala signaling and symptom reduction, observed in the current study, might relate to treatment-induced changes in the amygdala.

So far, one other study assessed the value of pre-treatment fMRI signals for predicting treatment-related changes (Perez et al., 2015). In this study, greater pre-treatment BOLD signal in the right anterior cingulate cortex during an emotional go/nogo task was associated with reduced improvement in terms of the factor 'constraint' of the multidimensional Personality Questionnaire. Moreover greater BOLD signal in the left posterior-medial OFC/ventral striatum was associated with reduced improvement in terms of the total score on the Affective Lability Scale. Together with these prior data, our findings strengthen the observation that greater responsiveness of limbic circuitry during affective action regulation renders BPD patients more resistant to clinical improvement after therapy. The differences between the findings of this study and ours might be exploited in future research into treatment-specific predictors of symptom reduction.

Note that the specificity of the current results to prediction of the success of DBT rather than other types of treatment remains unproven, given the absence of a control treatment. Thus, although the present results raise the question whether patients who exhibit enhanced amygdala responsiveness respond better to treatment regimes other than DBT (Schulze 2016), caution is warranted when interpreting our findings in terms of speaking selectively to DBT. We rather restrict our conclusions to the general case of clinical improvement, regardless of whether this improvement is elicited by DBT, attention and motivation, or time.

Three limitations of our study deserve special attention: First, our main result is based on a fairly small sample size. Although we assessed the robustness of the effect extensively, for example by cross validation (leave

one participant out procedure) and by permutation based analyses (SnPM), replication of our data is needed. Second, we used a 'real-life' BPD patient group with the majority of patients being on psychotropic medication and having multiple comorbidities. This choice of patient selection was at the expense of internal validity, which we deliberately traded off against enhanced external validity. The majority of patients in normal clinical practice with BPD has multiple comorbidities and, although discouraged in many guidelines, take psychotropic medications, such as selective serotonin inhibitors. Choosing for such a sample is in line with our ultimate aim to find useful biobehavioral markers to predict treatment success in clinical practice. Third, although we replicated the behavioral PIT results and part of the imaging results of our previous healthy control PIT study (Geurts et al., 2013a), we were not able to replicate the PIT-related brain-behavior correlations from this study. This might be due to insufficient power for replication (Button et al., 2013) of the current study, or depend on several differences between the study groups (e.g. gender). This should be addressed in future, larger replication studies and meta-analyses.

4.5 REFERENCES

ALLMAN MJ, DeLeon IG, Cataldo MF, Holland PC, Johnson AW (2010) Learning processes affecting human decision making: An assessment of reinforcer-selective Pavlovian-to-instrumental transfer following reinforcer devaluation. Journal of Experimental Psychology: Animal Behavior Processes 36:402–408.

AMERICAN psychiatric association T (2000) Diagnostic and statistical manual of mental disorders, Fourth edition, Text Revision. Washington, DC: American Psychiatric Association.

ARNTZ A, van den Hoorn M, Cornelis J, Verheul R, van den Bosch WMC, de Bie AJHT (2003) Reliability and validity of the borderline personality disorder severity index. J Pers Disord 17:45–59.

BALLEINE BW, Doherty JPOA (2009) Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. Neuropsychopharmacology 35:48–69.

BOLTON JM, Robinson J (2010) Population-attributable fractions of Axis I and Axis II mental disorders for suicide attempts: findings from a representative sample of the adult, noninstitutionalized US population. American Journal of Public Health, 100(12), 2473–2480.

BRETT M, Anton J-L, Valabregue R, Poline J-B (2002) Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain.

BUTTON KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14:365–376.

BÜCHEL C, Wise RJ, Mummery CJ, Poline JB, Friston KJ (1996) Nonlinear regression in parametric activation studies. NeuroImage 4:60–66.

CARDINAL RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience and Biobehavioral Reviews 26:321–352.

CORBIT LH (2005) Double Dissociation of Basolateral and Central Amygdala Lesions on the General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer. Journal of Neuroscience 25:962–970.

CORBIT LH, Balleine BW (2011) The General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer Are Differentially Mediated by the Nucleus Accumbens Core and Shell. Journal of Neuroscience 31:11786–11794.

ESTERMAN M, Tamber-Rosenau BJ, Chiu Y-C, Yantis S (2010) Avoiding non-independence in fMRI data analysis: Leave one subject out. NeuroImage 50:572–576.

GARBUSOW M, Schad DJ, Sebold M, Friedel E, Bernhardt N, Koch SP, et al. (2016). Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. Addiction Biology, 21(3), 719–731.

GEURTS DEM, Huys QJM, Ouden den HEM, Cools R (2013a) Aversive Pavlovian Control of Instrumental Behavior in Humans. Journal of Cognitive Neuroscience 25:1428–1441.

GEURTS DEM, Huys QJM, Ouden den HEM, Cools R (2013b) Serotonin and Aversive Pavlovian Control of Instrumental Behavior in Humans. Journal of Neuroscience 33:18932–18939.

GIESEN-BLOO J, Arntz A, Schouten E (n.d.) The borderline personality disorder checklist: psychometric evaluation and factorial structure in clinical and nonclinical samples.

GITELMAN DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. NeuroImage 19:200–207.

GOODMAN M, Carpenter D, Tang CY, Goldstein KE, Avedon J, Fernandez N, Mascitelli KA, Blair NJ, New AS, Triebwasser J, Siever LJ, Hazlett EA (2014) Journal of Psychiatric Research. Journal of psychiatric research, 57:108–116.

GRANT BF, Chou SP, Goldstein RB, Huang B, Stinson FS, Saha TD, Smith SM, Dawson DA, Pulay AJ, Pickering RP, Ruan WJ (2008) Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. J Clin Psychiatry 69:533–545.

GUNDERSON JG (2009) Borderline personality disorder: ontogeny of a diagnosis. Am J Psychiatry 166:530–539.

HEBART MN, Gläscher J (2015) Serotonin and dopamine differentially affect appetitive and aversive general Pavlovian-to-instrumental transfer. Psychopharmacologia, 232(2), 437–451.

HEINZ A, Schlagenhauf F, Beck A, Wackerhagen C (2016) Dimensional psychiatry: mental disorders as dysfunctions of basic learning mechanisms. Journal of Neural Transmission, 123(8), 809–821.

HOERTEL N, López S, Wang S, González-Pinto A, Limosin F, Blanco C (2015) Generalizability of pharmacological and psychotherapy clinical trial results for borderline personality disorder to community samples. Personality Disorders: Theory, Research, and Treatment 6:81–87.

HUYS QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P (2011) Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding Rangel A, ed. PLoS Comput Biol 7:e1002028.

HUYS QJM, Gölzer M, Friedel E, Heinz A, Cools R, Dayan P, Dolan RJ (2016) The specificity of Pavlovian regulation is associated with recovery from depression. Psychol Med:1–9.

JACOB G, Zvonik K, Kamphausen S, Sebastian A, Maier S, Philipsen A, Tebartz van Elst L, Lieb K, Tüscher O (2013) Emotional modulation of motor response inhibition in women with borderline personality disorder: an fMRI study. J Psychiatry Neurosci 38:164–172.

JONES KA, Menniti FS, Sivarao DV (2015) Translational psychiatry-light at the end of the tunnel. Annals of the New York Academy of Sciences 1344:1–11.

KRAUSE-UTZ A, Winter D, Niedtfeld I, Schmahl C (2014) The Latest Neuroimaging Findings in Borderline Personality Disorder. Curr Psychiatry Rep 16:438.

LANA F, Fernández-San Martín MI (2013) To what extent are specific psychotherapies for borderline personality disorders efficacious? A systematic review of published randomised controlled trials. Actas Esp Psiquiatr 41:242–252.

LARSON CL, Baskin-Sommers AR, Stout DM, Balderston NL, Curtin JJ, Schultz DH, Kiehl KA, Newman JP (2013) The interplay of attention and emotion: top-down attention modulates amygdala activation in psychopathy. Cogn Affect Behav Neurosci 13:757–770.

LEWIS AH, Niznikiewicz MA, Delamater AR, Delgado MR (2013) Avoidance-based human Pavlovian-to-instrumental transfer. Eur J Neurosci.

LINEHAN MM (1993) Cognitive-behavioral Treatment of Borderline Personality disorder. New York: NY Guilford Press.

LY V, Cools R, Roelofs K (2014) Aversive disinhibition of behavior and striatal signaling in social avoidance. Social Cognitive and Affective Neuroscience 9:1530–1536.

MCLAREN DG, Ries ML, Xu G, Johnson SC (2012) A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. NeuroImage 61:1277–1286.

PEREZ DL, Vago DR, Pan H, Root J, Tuescher O, Fuchs BH, Leung L, Epstein J, Cain NM, Clarkin JF, Lenzenweger MF, Kernberg OF, Levy KN, Silbersweig DA, Stern E (2015) Frontolimbic Neural Circuit Changes in Emotional Processing and Inhibitory Control Associated With Clinical Improvement Following Transference-Focused Psychotherapy in Borderline Personality Disorder. Psychiatry Clin Neurosci.

POOL E, Brosch T, Delplanque S (2015) Stress increases cue-triggered "wanting" for sweet reward in humans. Journal of Experimental Psychology. Animal Learning and Cognition, 41(2), 128–136.

POSER BA, Versluis MJ, Hoogduin JM, Norris DG (2006) BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. Magn Reson Med 55:1227–1235.

PREVOST C, Liljeholm M, Tyszka JM, O'Doherty JP (2012) Neural Correlates of Specific and General Pavlovian-to-Instrumental Transfer within Human Amygdalar Subregions: A High-Resolution fMRI Study. Journal of Neuroscience 32:8383–8390.

QUAIL SL, Morris RW, Balleine BW (2016) Stress associated changes in Pavlovian-instrumental transfer in humans. The Quarterly Journal of Experimental Psychology:1–11.

ROSENTHAL MZ, Gratz KL, Kosson DS, Cheavens JS, Lejuez CW, Lynch TR (2008) Borderline personality disorder and emotional responding: A review of the research literature. Clinical Psychology Review 28:75–91.

SALVADOR R, Vega D, Pascual JC, Marco J, Canales-Rodríguez EJ, Aguilar S, Anguera M, Soto A, Ribas J, Soler J, Maristany T, Rodríguez-Fornells A, Pomarol-Clotet E (2016) Archival Report. BPS 79:107–116.

SCHNELL K, Herpertz SC (2007) Effects of dialectic-behavioral-therapy on the neural correlates of affective hyperarousal in borderline personality disorder. Journal of Psychiatric Research 41:837–847.

SCHULZE L, Schmahl C, Niedtfeld I (2016) Archival Report. BPS 79:97–106.

SELBY EA, Anestis MD, Bender TW, Joiner TE (2009) An exploration of the emotional cascade model in borderline personality disorder. Journal of Abnormal Psychology 118:375–387.

SILBERSWEIG D, Clarkin JF, Goldstein M, Kernberg OF, Tuescher O, Levy KN, Brendel G, Pan H, Beutel M, Pavony MT, Epstein J, Lenzenweger MF, Thomas KM, Posner MI, Stern E (2007) Failure of frontolimbic inhibitory function in the context of negative emotion in borderline personality disorder. Am J Psychiatry 164:1832–1841.

SOLOFF PH, White R, Omari A, Ramaseshan K, Diwadkar VA (2015) Psychiatry Research: Neuroimaging. Psychiatry Research: Neuroimaging 233:23–35.

STOFFERS JM, Völlm BA, Rücker G, Timmer A, Huband N, Lieb K (2012) Psychological therapies for people with borderline personality disorder. Cochrane Database Syst Rev 8:CD005652.

TALMI D, Seymour B, Dayan P, Dolan RJ (2008) Human Pavlovian Instrumental Transfer. Journal of Neuroscience 28:360–368.

TZOURIO-MAZOYER N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. NeuroImage 15:273–289.

VALENTIN VV, Dickinson A, O'Doherty JP (2007) Determining the Neural Substrates of Goal-Directed Learning in the Human Brain. Journal of Neuroscience 27:4019–4026.

WATSON P, Wiers RW, Hommel B, de Wit S (2014) Working for food you don't desire. Cues interfere with goal-directedfood-seeking. Appetite 79:139–148.

WUNDERLICH K, Rangel A, O'Doherty JP (2009) Neural computations underlying action-based decision making in the human brain. Proceedings of the National Academy of Sciences 106:17199–17204.

WUNSCH E-M, Kliem S, Kröger C (2014) Behaviour Research and Therapy. Behaviour Research and Therapy 60:1–7.

4.6 SUPPLEMENT

4.6.1 *Supplementary materials and methods*

Additional details on subject inclusion

Aversive PIT and symptom reduction Exclusion criteria for all participants were severe somatic and neurological illness, mental retardation, severe hearing and visual disabilities, illicit drug use in the week before the experiment, alcohol or benzodiazepine use within 24-hours of the experiment and smoking 3 hours before scanning: Twenty-three patients volunteered. One patient could not participate due to metal in his spine, and one patient failed to attend the screening session. Twenty-one patients attended the intake session. For subsequent assessments, two patients were excluded because they did not meet the DSM-IV criteria of BPD. One patient reported not wanting to participate out of anxiety for the aversive juice. Two patients did not show up on the first scan session and we were unable to subsequently establish contact. During the first scan session one patient experienced a panic attack. The final dataset included 15 patients. Furthermore, one subject did not attend the one year follow-up session. Four of the other 14 patients who attended this session reported not having completed the full year of therapy.

Image acquisition

ME-EPI sequence details: 38 axial-oblique slices, repetition time, 2.250s; echo-times: 9.7, 20.3, 31, 41 and 52ms; in plane resolution, 3.5x3.5mm; slice thickness, 2.5mm; distance factor 0.17; flip angle, 96. Visual stimuli were projected on a screen and viewed through a mirror attached to the head coil. In addition, a high-resolution T1-weighted magnetization-prepared rapid-acquisition gradient echo anatomical scan was obtained from each subject (192 sagittal slices; repetition time, 2.3s; echo time, 3.03ms; voxel size 1.0 x 1.0 x 1.0 mm; field of view 256 mm).

Supplementary analyses of instrumental and Pavlovian training

The behavioural data were analysed using the statistic software SPSS 16.0. First, we assessed whether subjects learnt the instrumental training task. The proportion of correct responses was calculated for the first ten and last ten trials separately for each of the four trial types. Performance (p(correct), not normally distributed) was compared between groups and between the beginning and the end of instrumental training by means of Wilcoxon Signed Rank and Mann Whitney U tests respectively. To assess differences in Pavlovian conditioning between the groups we compared performance on the Pavlovian query trials (p(correct) , not normally distributed) by means of Mann Whitney U tests. In addition, liking ratings (not normally distributed) of the CSs before and after the experiment were analyzed using Mann Whitney U and Wilcoxon Signed Rank tests.

Supplementary neuroimaging analyses

Secondary to our main analysis we fully explored the 2x2x2 rmANOVA using the parametric PIT regressors with Action Context (approach/withdrawal) and Valence (neutral/aversive) and Group(healthy controls/borderline personality disorder).

Moreover we additionally assessed the relation between CS dependent BOLD signal change during the PIT stage and average behavioural PIT scores on a subject by subject basis. Therefore, in addition to the main effect of Valence [(approach neutral+withdrawal neutral) - (approach aversive+withdrawal aversive)] (see main paper), we calculated an interaction between Valence and Action Context [(approach neutral-approach aversive)-(withdrawal neutral-withdrawal aversive)]. The resulting SPMs for each contrast were then used in a two-sample t-test at the group-level with behavioural aversive PIT-effects (p(go) (Geurts et al., 2013a; 2013b) and average number of button presses (Geurts et al., 2013a)) as a covariate for each group separately enabling comparison between groups. These analyses revealed additional regions in which individual differences in BOLD responses were linearly associated with individual differences in behavioural PIT in terms of choice and vigour respectively across and between groups.

Generalized psychophysiological interaction (gPPI) analyses

We used a generalized form of context-dependent psychophysiological interaction (gPPI, http://brainmap.wisc.edu/PPI, McLaren et al., 2012). To compose the physiological variable, the extracted mean time series of the

BOLD signal from the vmPFC blob (Figure 4.3, main paper) were temporally filtered, mean corrected, and de-convolved to generate the time series of the neural signal for the vmPFC for each individual subject. These time series of neural signal were then multiplied by the onset times of the trials with different CS Valence and separately with the parametric PIT-regressor (vector consisting of total number of button presses per trial per Action Context and CS Valence). The products were then re-convolved with the canonical HRF to obtain the interaction term or PPI variable (Gitelman et al., 2003). Next, these regressors were added to the first level GLM described in the main paper. The parameter estimates of the PIT-related PPI regressors (those based on the product of the seed time series and parametric PIT-regressor) quantify the relation with trial-by-trial instrumental action (i.e. number of button presses) and functional connectivity between the seed region (i.e. vmPFC) and other regions. Contrasting the PIT-related PPI regressors between CS Valence (and/or Action Context) reveals regions which functional connectivity with the vmPFC is differentially related to instrumental action as a function of CS Valence, thus representing PIT-related functional connectivity.

Based on Geurts et al. (2013a) we expected that functional connectivity between the vmPFC and caudate nucleus would be dependent on CS Valence during withdrawal. Therefore we submitted the beta-estimates resulting from the contrast [neutral | withdrawal] – [aversive | withdrawal] to a two sample t-test. We also tested, following the results of Geurts et al. (Geurts et al., 2013a) whether the beta-estimates resulting from the contrast [neutral | approach] – [aversive | approach] were related to behavioral aversive PIT in terms of button presses [BP | approach&neutral - BP | approach&aversive].

4.6.2 *Supplementary Results*

Instrumental conditioning

Overall subjects learned to make correct choices during the instrumental learning stage indicated by an increasing number of correct responses over time (p(correct | Time Bin1) vs p(correct | Time Bin2): Related Samples Wilcoxon Signed Rank Test: $p<.001$). There was no difference in learning between the groups, not across Action Contexts and not for one of Action Contexts separately (Mann Whitney U test: all $p>.616$). Furthermore, during the PIT stage performance was the same for the groups across both Action Contexts and for each Context separately (Mann Whitney U test: all $p>.616$).

Pavlovian conditioning

Mann-Whitney U tests showed there were no differences between the groups in how they rated the aversive and appetitive juice before and after conditioning (for all 4 comparisons p>.381). There was also no difference between the groups in pre to post changes in rating for the different juices (p>.669). Furthermore, there was no difference in performance on the Pavlovian query trials (Mann-Whitney U test: mean proportion correct over blocks in BPD: 94%; SEM: 2.2; range: 75-100%; HC: 93%; SEM: 3.4; range: 50-100%, p=.800). VAS ratings for the Pavlovian CSs showed that the aversive CS became aversive to the participants (Wilcoxon Signed Rank Test: p=.001, one-tailed) and that the neutral and appetitive CSs did not change (Wilcoxon Signed Rank Test: p>.181, one-tailed; Figure S1). However after conditioning the aversive CS did not differ significantly on VAS rating from the neutral CS, but did differ from the appetitive CS (Mann-Whitney U test: p=.031, one-tailed) and the appetitive CS was judged more appetitive than the neutral CS (Wilcoxon Signed Rank Test: p=.027, one-tailed). None of the VAS ratings for the Pavlovian CSs or their changes from before to after conditioning differed between the groups (All Mann-Whithney U tests: p>.381).

Pavlovian to instrumental transfer stage

VIGOUR Analysis of vigour (i.e. number of button presses) revealed a main effect of Action Context, due to more vigorous responding during approach than withdrawal (Supplementary table 4.3, $F_{(1,29)}$=33.7, p<.001). There were no significant differences between the groups in terms of the vigour of responding.

APPETITIVE PIT As in previous studies with this task, we did not observe significant appetitive PIT (neutral vs. appetitive). The rmANOVA with Action Context(approach/withdrawal) and CS Valence(neutral/appetitive) as within subject factors and Group(HC/BPD) as between subject factor did not reveal any significant PIT-effects either with choice (p(go), all F<2.4, all p>.05) or with vigour (number of button presses, all F<3.18, all p>.05) as dependent variable.

Supplementary neuroimaging results

In addition to the action specific signal in the vmPFC (see main paper), we also observed action-specific signal in the precuneus, lingual and middle occipital gyrus (Supplementary Table 4.4, cf. Geurts et al., 2013a).

Significant brain-behaviour correlations were observed in the left amygdala (Table S3). Subjects showing increased Action Context specific aversive PIT in terms of behavior also showed Action Context specific responses of the amygdala to the aversive compared to the neutral CS. Thus, subjects who showed increased aversive inhibition of approach actions together with increased aversive activation of withdrawal actions also showed this aversive CS induced pattern with respect to amydgalar BOLD response.

No significant findings were revealed by the additional analysis of functional connectivity between the vmPFC and caudate nucleus.

4.6.3 *Supplementary Tables*

SJ#	Psychoactive Medication (prescribed and non-prescribed)	SCID-II BPD #items	MINI -plus classifications @baseline	Therapy Completed
1		6	-	Yes
2	topiramate	7	MDD, PTSD, alcohol dependence and abuse, boulemia nervosa, hypochondria	Yes
3	citalopram	8	-	No
4	citalopram	6	Past abuse and dependence of marihuana and XTC	No
5	ventolin	7	MDD	No
6	melatonin	8	MDD, Agorafobia	Yes
7	lamotrigine, trazolan, ezomeprazol, zeracette, diazepam	8	Agorafobia, PTSD, Alcohol dependence and abuse, GAD, ADHD	Yes
8	diazepam	8	-	No
9	oxazepam	6	Bipolar II disorder	Yes
10	-	8	Abuse of XTC and GHB, PMS	Yes
11	-	8	MDD, panic disorder, agoraphobia, social fobia	Yes
12	topiramate	9	MDD, PTSD,	Yes
13	paroxetine	6	Dysthymia, simple fobia, boulimea nervosa	Yes
14	-	9	MDD, agorafobia, social fobia, PTSD, ADHD	Yes
15	jasmin OAC, simbicort, oxycontin, temazepam	8	MDD, Boulimea nervosa	No

Table 4.2 *– Medication use and MINI-plus classifications before treatment. Abbreviations: MDD, major depressive disorder; PTSD, post traumatic stress disorder; ADHD, attention deficit and hyperactivity disorder; PMS, post-menstrual syndrome; GAD: general anxiety disorder*

	Action Context			
	Approach		Withdrawal	
	HC	BPD	HC	BPD
Appetitive	7.1(1.6)	7.8(1.3)	8.1(1.3)	8.4(1.2)
Neutral	6.8(1.6)	7.7(1.2)	8.0(1.5)	8.3(1.3)
Aversive	6.9(1.6)	7.5(1.8)	8.1(1.2)	8.6(1.2)

Table 4.3 *– Presented are the average number of button presses for the healthy control (HC) and borderline personality disorder (BPD) group as a function of Action Context (approach/withdrawal) and CS Valence (appetitive/neutral/aversive) during the Pavlovian-instrumental transfer stage (standard deviation)).*

Location	**k**	**X**	**Y**	**Z**	**Z-value**	**P-value FWE- cor**
Full Factorial: Group x Action Context x Valence (PIT-regressor)						
F-test: Main effect of Action Context:						
Linugal and Calcarine (Bil)	3547	-12	-76	2	Inf	P_{WB}<.001
Mid Occipital (L)	147	-42	-68	0	5.6	P_{WB}<.001
Supramarginal gyrus (L)	341	-60	-22	38	5.95	P_{WB}<.001
Mid Occiptal (R)	111	36	-82	8	4.96	P_{WB}=.025
Precentral (L)	78	-22	-12	54	4.81	P_{WB}=.048
F-test: Interaction Group x Action Context x Valence						
White matter near inferior frontal gyrus	60	32	-2	26	4.88	P_{WB}=.036
2 sample T-test: Main regressor: Action Context specific aversive PIT contrast + covariate of interest: behavioural action specific aversive PIT (button presses) [Neu | Approach]-[Ave | Appraoch] – ([Neu | Withdrawal]-[Ave | Withdrawal])						
T-test: main positive effect of covariate:						
Amygdala left	45	-30	0	-18	4.86	P_{SV}=.003

Table 4.4 *– Supplementary fMRI results*

5

PSYCHOPATHY SEVERITY IS ASSOCIATED WITH REDUCED AVERSIVE PAVLOVIAN TO INSTRUMENTAL TRANSFER

D.E.M. Geurts K. von Borries, Q.J.M. Huys, B.H. Bulten, R.J. Verkes, R. Cools

In preparation

Background: Psychopathy is characterized by cold instrumental (aggressive) behavior and has been associated with abnormal aversive processing. However, the consequences of such abnormal aversive processing for instrumental action and associated neural mechanisms are unclear.
Methods: Here we address this issue by using event-related functional magnetic resonance imaging in 15 psychopathic criminals and 18 matched controls during the performance of an aversive Pavlovian-to-instrumental transfer paradigm. This paradigm allowed us to assess the degree to which aversive Pavlovian cues inhibit instrumental action.
Results: Clinical psychopathy severity correlated with an attenuation of aversive Pavlovian inhibition of instrumental action. Moreover, there was an anomalous positive association between aversive inhibition of action and aversive inhibition of BOLD signal in the caudate nucleus in psychopathic criminals.
Conclusions: These results show that clinical psychopathic severity is associated with reduced transfer of aversive Pavlovian cues to instrumental action inhibition. These findings demonstrate that psychopathy involves abnormal impact of aversive processing on instrumental behavior and raise the hypothesis that instrumental aggression reflects a lack of aversive Pavlovian inhibition. This aversive Pavlovian inhibition might be due to inappropriate transfer of aversive Pavlovian values to neural systems involving the caudate nucleus.

5.1 INTRODUCTION

Instrumental decision making is well known to be susceptible to emotional/affective influences(Estes and Skinner, 1941; Damasio, 1997). Evidence suggests that this affective biasing of action selection reflects an interaction between distinct behavioral control systems(Cardinal et al., 2002; Dayan et al., 2006; Kahneman and Frederick, 2007). For example, instrumentally controlled action selection is biased by a Pavlovian or 'affective' system that regulates innately specified responses to aversive stimuli(Dayan and Seymour, 2008). Critically, anomalies in the interaction between these Pavlovian and instrumental control systems have been proposed to account for behavioural impairments seen in a wide variety of neuropsychiatric disorders (Dayan et al., 2006; Seymour and Dolan, 2008; Heinz et al., 2016).

Psychopathy is one such disorder, being associated with cold instrumental (aggressive) behavior, and imposing a large burden on individual victims and society as a whole. Psychopathy is characterized by a complex of affective and behavioral anomalies(Hare, 2003) and psychopathic criminals (PCs) are prone to commit violent crimes(Porter and Woodworth, 2006) with high rates of recidivism even after prison sentences(Serin and Amos, 1995). People fulfilling the criteria for psychopathy are overrepresented in the US prison population: about 25% of inmates are diagnosed with psychopathy compared with 1% of the general population(Hare, 2003; Porter and Woodworth, 2006).

A core feature of psychopathic criminality is the instrumental nature of the crimes committed(Blair, 2001). These crimes are premeditated and committed to achieve a desired goal at the expense of others. Despite the centrality of cold instrumental action in clinical observations and in elaborate cognitive models of psychopathy (e.g. the violence inhibition model (Blair, 2005)), neuroscientific research on the mechanisms of instrumental action in the face of aversive cues is scarce.

So far neurobehavioral research on psychopathy has focused mainly on attenuated affective (primarily aversive) processing per se and its underlying neural circuitry(Brook et al., 2013). There is consistent evidence (albeit in small samples) that PCs respond normally to unconditioned aversive Pavlovian stimuli (US), but that their psychophysiological responses to conditioned aversive stimuli (CS) are compromised(Flor et al., 2002; Veit et al., 2002; Birbaumer et al., 2005; Rothemund et al., 2012). However it is unclear how such a deficiency in aversive information processing is related

to the broad behavioural phenotype of psychopathy. Studies focusing on affective anomalies in itself do not provide direct insight in the behavioral deficits that might stem from these affective anomalies. Here we consider how instrumental action is altered by aversive affective cues, thus directly addressing behaviour. Specifically we investigate the impact of aversive affective information processing on instrumental action and the associated neural mechanisms in PCs.

We focus on conditioned suppression, where an aversive Pavlovian cue inhibits ongoing instrumental action. Following prior work(Huys et al., 2011; Geurts et al., 2013a; 2013b), we refer to this as aversive Pavlovian to instrumental transfer (PIT). We hypothesize that PCs exhibit reduced aversive Pavlovian inhibition also known as conditioned suppression. This hypothesis concurs generally with our previous observation that aversive PIT is attenuated by serotonin depletion in healthy controls(Geurts et al., 2013b). This sensitivity to serotonin manipulation is pertinent in the current context, given prior observations that psychopathy severity and aggression have been associated with reduced serotonin transmission ((Coccaro, 1992; Soderstrom et al., 2001; 2003), but see(Yildirim and Derksen, 2013)).

We also aimed to assess the neural mechanisms underlying the aversive PIT effects. Animal and human studies consistently implicate fronto-striatal brain regions in instrumental action, especially the dorsomedial (caudate nucleus) and dorsolateral (putamen) parts of the striatum and the ventromedial regions of the prefrontal cortex(Valentin et al., 2007; Balleine and Doherty, 2009; Tricomi et al., 2009; Wunderlich et al., 2009; Dolan and Dayan, 2013). Furthermore, we showed that aversive PIT was accompanied by modulation of functional fronto-striatal connectivity, specifically between the ventromedial prefrontal cortex and caudate nucleus(Geurts et al., 2013a). In addition, affectiveversive information is known to influence instrumental actions via the amygdala(Cardinal et al., 2002; Talmi et al., 2008; Balleine and Doherty, 2009; Prevost et al., 2012; Geurts et al., 2013a; Ly et al., 2014), and extensive evidence implicates dysfunction of the amygdala in psychopathy(Veit et al., 2002; Birbaumer et al., 2005; Blair, 2008; Glenn and Raine, 2009; e.g. Moul et al., 2012). Thus, we anticipated, first, that psychopathy would be accompanied by reduced aversive PIT; and would be reduced in psychopathy.; ssecond, that psychopathy would be accompanied by differential modulation of frontal and striatal brain regions, especially by regions processing aversive information such as the amygdala; and third that these two findings would be related. To this end, we focused our primary analyses of transfer on the striatum, ventromedial prefrontal cortex and the amygdala.

Table 5.1 *– Group characteristics (mean, standard deviation) of the group of psychopathic criminals (PP) and healthy matched control subjects*

	PP (n=15)	HC (n=18)	Statistics (p-value)
Age	40.2 (9.1)	41.2 (10.4)	.78
IQ (NLV)	101.7 (8.8)	101.5 (8.7)	.96
PCL-R total	30.7 (4.0)	-	-
PCL-R factor 1	11.9 (2.9)	-	-
PCL-R factor 2	13.9 (2.1)	-	-

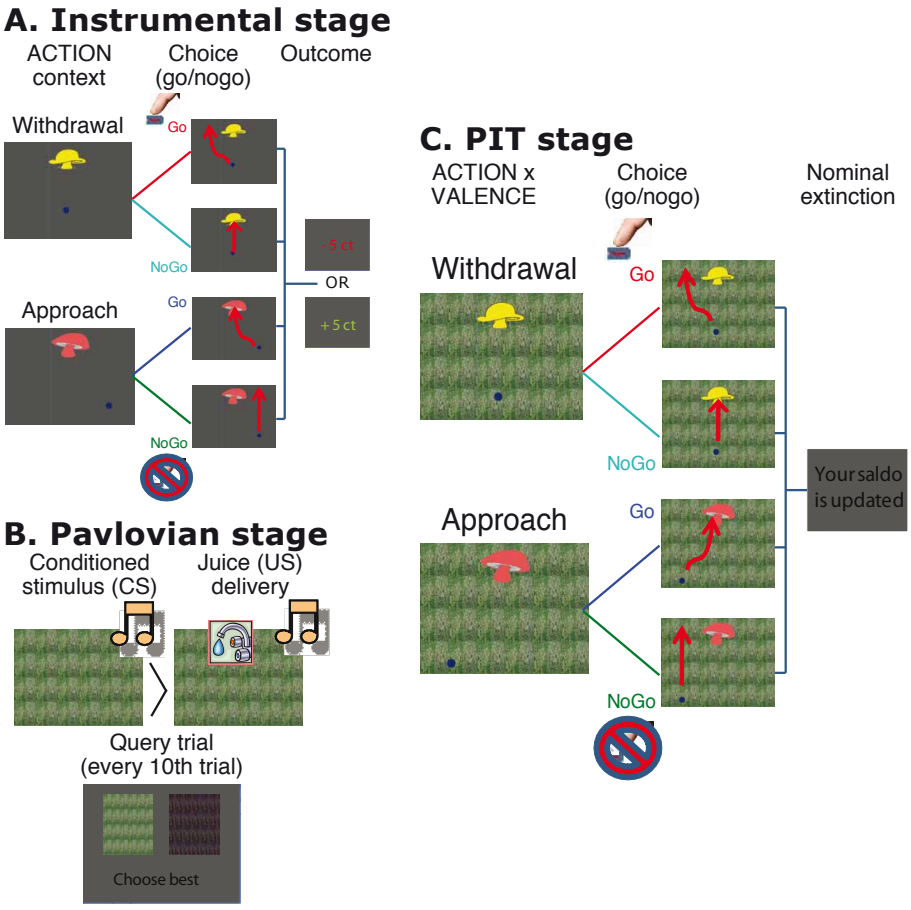
Exclusion criteria for both groups were: (i) Use of alcohol more than 3 units/day during the week preceding the experimental measure and use of alcohol within 24 hours of the measurement; (ii) Use of cannabis or other illicit drugs within the week before measurement and use of psychotropic medication other than oxazepam during the 5 days before measurement; (iii) Use of oxazepam within 12 hours before measurement; (iv) Smoking within 3 hours before measurement; (v) History of trauma capitis, visual and auditory disorders, neurological disorders, first degree relative with any relevant neurological disorders.

5.2 METHODS AND MATERIALS

5.2.1 *Participants*

Eighteen male PCs (3 left-handed) volunteered and were selected based on available information about clinical status and history from an in-patient population of a forensic hospital (Suplementary Material and Methods). The PCs were diagnosed with a score of $\geq$ 26 on the Hare Psychopathy Check List-Revised (PCL-R(Hare, 2003), Table 5.1). Additionally twenty healthy men matched for age and IQ without criminal records or a history of psychiatric disorders were recruited from among the employees of the same hospital by advertisement. Participants in both groups were screened for drug use and for medical/neurological history (Suplementary Material and Methods, Table 5.1). The local Medical Ethical Committee (NL30545.091.09) approved the study.

Two PCs withdrew from participation and 1 PC was excluded because of excessive head movement (> 2x voxel size). Two healthy controls (HCs) were excluded because their behavioral data suggested they did not follow the instructions during the PIT stage (despite instructions to play the instrumental game (see paradigm) these participants determined their ac-



tions solely on the Pavlovian CS, but never to the instrumental stimuli in more than half of the trials: 58% and 83% resp., compared to on average 1%, range 0-17%, for all other participants).

Moreover, due to technical reasons and excessive head movement only one of two runs could be analyzed for one HC and two PCs. Thus, we analyzed datasets of 15 PCs and 18 HCs.

Figure 5.1 *– Task details.* ***A. Instrumental stage.*** *Trials started with the appearance of the instrumental stimulus at the top center of the screen and of a dot at the bottom of the screen. In approach trials, the dot started either on the left or on the right bottom side of the screen. Participants could choose to do nothing (approach-no-go), in which case the dot would wiggle past the instrumental stimulus. Alternatively, they could push the button repeatedly to steer the dot through the instrumental stimulus (approach-go). In withdrawal trials, the dot started centrally at the bottom beneath the instrumental stimulus. Participants could choose to push the button repeatedly to avoid moving through instrumental stimulus (withdrawal-go) or to do nothing (withdrawal-no-go). The four possible trajectories are drawn in the figure (red lines). If the dot entered the goal region, then the instrumental stimulus was collected. After the dot moved outside the window feedback was provided. Thus, there were 2 Action contexts (approach and withdrawal), with each 4 different instrumental stimuli, with 2 stimuli resulting more often in reward after a go-action and 2 resulting more often in reward after a no-go. Each stimulus was presented 10 times, resulting in (2x4x10=) 80 instrumental trials (divided in miniblocks of 8 withdrawal or approach trials). The straight line just to one side of the instrumental stimulus could not be crossed by the dot. Timings were as follows: Instrumental stimuli were presented for 2.5 sec, during which responses were collected. After 2.5 sec, feedback was presented for 1 sec. The intertrial interval (ITI) was 1 sec (blank screen).* ***B. Pavlovian stage.*** *Each Pavlovian CS was presented 20 times, and for each session there was a separate set of three stimuli. Stimulus presentation order was fully randomized across participants. Stimulus duration was 4.5 sec, and juice delivery (2ml) occurred between 0 and 1.5 sec after stimulus onset. The ITI was 1 sec. Query trials were presented after every 10 Pavlovian trials. On these trials, participants chose one of the two presented Pavlovian (audiovisual)_stimuli (presented for 2 sec; ITI 0.5 sec) without any feedback.* ***C. PIT stage.*** *The PIT stage paralleled the instrumental training, except that Pavlovian CSs tiled the background. Each instrumental stimulus was presented 12 times and each Pavlovian CS 32 times counterbalanced over the different instrumental stimuli. No outcomes were presented, but participants were instructed that their choices counted toward the final total. Participants were explicitly instructed that the juices were collected outside the scanner, and they agreed before the start of the experiment to drink them afterward. Timing of one trial was as follows: 250 msec after the onset of the Pavlovian stimulus, the instrumental stimulus (and dot) was overlaid on top of this Pavlovian stimulus. Duration of the instrumental stimulus was 2.5 sec; duration of the Pavlovian stimulus was 2.75 sec. Upon offset of both stimuli, feedback was presented, which consisted only of the words "Balance is updated" (duration = 1 sec, ITI = 1 sec). Note that there were two runs in which all three stages (with new independent Pavlovian and instrumental stimuli) were assessed.*

5.2.2 *Pavlovian-instrumental transfer paradigm*

Subjects performed a computerized task to assess aversive PIT(Geurts et al., 2013a, Figure 5.1). The experiment consisted of three stages: (1) instrumental, (2) Pavlovian and (3) PIT stage. The instrumental stage contained two Action Contexts: (i) a context in which the active response led to an approach action and (ii) another in which the active response led to a withdrawal action. In each context 2 go-stimuli and 2 nogo-stimuli were repeatedly presented to the participant (Figure 5.1A). In the approach Action Context subjects learned through monetary feedback (wins and losses) whether to 'collect' the instrumental stimulus (approach-go) or not (approach-no-go). In the withdrawal Action Context they learned to avoid collecting instrumental stimuli (withdrawal-go) or not (withdrawal-no-go). Instrumental stimuli were randomly assigned to one of the four trial types. Thus, in both the approach and withdrawal Action Contexts, there were 2 go-stimuli, which yielded reward more often (i.e. ~85% of the cases) after active responses (and punishment after not responding), and 2 nogo-stimuli, which yielded reward more often (i.e. also ~85% of the cases) after not responding (and punishment after go-responding).

The second, Pavlovian stage consisted of repeated presentation of three audiovisual stimuli (Figure 5.1B): The appetitive and aversive conditioned stimuli (CS) were followed, respectively, by appetitive or aversive juice (i.e. the unconditioned stimuli USs) on 50% of trials. The neutral CS resulted in no outcome. The appetitive juice was based on subjective preference for apple, orange or strawberry lemonade. The aversive juice was a bitter magnesium sulphate solution (0.3M). Conditioning was assessed in two ways: (1) subjects indicated the degree to which they liked each of the CSs (and USs) by use of visual analogue scales (VAS), before and after the experiment; (2) subjects chose one of the two presented Pavlovian stimuli (presented for 2s; ITI 0.5s) in extinction on 12 interspersed query trials.

In the third, PIT, stage stimulus presentation was the same as in the instrumental stage, except that Pavlovian stimuli tiled the background from 250ms before (Larson et al., 2013) and during the instrumental trial, and no outcomes were presented (Figure 5.1C). Subjects were instructed that their choices counted towards the final monetary total, and that the juices associated with the Pavlovian outcomes were collected outside the scanner for them to drink afterwards. There were 2 independent (i.e. other stimuli/CSs) runs including all three stages separated by a 2 minute break.

5.2.3 *Image acquisition*

Whole-brain imaging was performed on a 3 Tesla MR scanner (Magnetrom Trio Tim, Siemens Medical Systems, Erlangen, Germany). Functional data were obtained using a multi-echo gradient T2*-weighted echo-planar scanning sequence(Poser et al., 2006)(Supplementary Materials).

5.2.4 *Behavioural data analysis*

The behavioural data were analyzed using the statistic software SPSS 16.0 and Matlab 2009b.

Instrumental training

First, the proportion of correct responses was calculated for the first ten and last ten trials for each of the 4 trial types (covering all 80 instrumental trials). To assess whether subjects learned to make the correct choice, data were averaged across sessions and submitted to a repeated measures analysis of variance (rmANOVA) with Time (beginning/end of instrumental training), Action Context(approach/withdrawal) and Response(go/nogo) as within-subject and Group(HCs/PCs) as between-subject factor. Second, we assessed whether the learned behaviour generalized to the PIT stage. Therefore the factor Time was changed to include 3 levels: the end of the instrumental training and the beginning and the end of the PIT stage.

Pavlovian conditioning

Nonparametric tests were used to assess the proportion of correct responses on Pavlovian query trials and pre- and post conditioning VAS ratings of the CS, because data were not distributed normally.

Pavlovian-instrumental transfer

The behavioural outcome measures were proportion of go actions (i.e. p(go)) and the number of button presses on go-trials as a function of trial type (i.e. Action Context and CS Valence). P(go) provides the opportunity to assess PIT effects i.e. the influence of CS Valence and Action Context on choice (i.e. whether to go or not to go), whereas the number of button presses on go-trials indicates transfer effects on vigour. Both dependent variables were averaged across runs before they were submitted

to rmANOVAs with Action Context(approach/withdrawal), and CS Valence(neutral/aversive) as within-subject factors and Group(HCs/PCs) as a between-subject factor. Note that we focused our analyses on aversive PIT, based on our hypothesis (see introduction) and on our previous work (n=33) showing that the current paradigm was not sensitive to (and therefore not valid to assess) appetitive PIT (neutral vs. appetitive, (Geurts et al., 2013a),Supplementary Results). The PCL-R-score was added as a covariate to assess its association with aversive PIT.

5.2.5 *fMRI analysis*

fMRI analysis was performed with SPM5 software (Wellcome Trust Centre for Cognitive Neuroimaging, London, UK). Pre-processing steps and first-level fMRI analysis were exactly as described by Geurts(2013a): Preprocessing steps included applying a PAID-weight algorithm(Poser et al., 2006) to combine the different echoes, slice-time correction, coregistration, normalization based on parameters estimated through segmentation of the structural images, and smoothing.

The primary analysis was restricted to the PIT-stage. At the subject level a general linear model (GLM) was specified with 6 main regressors (4 of interest) representing the onset of the six different PIT trials of this paradigm (Action Context(approach/withdrawal) x Valence(appetitive/neutral/aversive)). For each main regressor two additional parametric regressors were added(Büchel et al., 1996): The PIT-regressor(Talmi et al., 2008) was a parametric modulator of BOLD responses by the number of button presses per trial. A further parametric regressor contained the expectation associated with each instrumental stimulus (the Q-value) per trial as estimated from a model-based analysis of behaviour(Huys et al., 2011). This was done based on prior data showing that BOLD signal in the prefrontal cortex and striatum, our regions of interest, covary with instrumental action value(Valentin et al., 2007; Wunderlich et al., 2009). As such, this approach maximized the degree to which our GLM captured variability in relevant BOLD signal. Furthermore, realignment parameters were added, high-pass filtering (128s) was applied and parameter estimates were obtained by maximum-likelihood estimation (AR1).

The parameter estimates for the 4 parametric PIT-regressors were used in a 2x2x2 rmANOVA at the group-level (with random effects) with Action Context(approach/withdrawal) and Valence (neutral/aversive) as within-subject factors and GROUP (healthy controls/PCs) as between-subjects factor. Planned contrasts were the same as in Geurts et al.(Geurts et al., 2013a),

but now assessed as a function of Group: [neutral-aversive] to reveal regions involved in aversive PIT across Action Contexts, and [(approach neutral-approach aversive) - (withdrawal neutral-withdrawal aversive)] to reveal regions involved in action-specific aversive PIT, and [approach–withdrawal] to reveal action-specific regions

To capture additional PIT signal that is related to stable patterns of behavior beyond trial-by-trial variation in instrumental vigour we contrasted the main regressors(Talmi et al., 2008; Geurts et al., 2013a) at the subject-level to calculate the main effect of Valence [(approach neutral+withdrawal neutral) - (approach aversive+withdrawal aversive)] and an interaction between Valence and Action Context [(approach neutral-approach aversive)-(withdrawal neutral-withdrawal aversive)]. The resulting SPMs for each contrast were then used in a two-sample t-test at the group-level with behavioural aversive PIT-effects (p(go)(Geurts et al., 2013b) and average number of button presses(Geurts et al., 2013a)) as a covariate for each group separately enabling comparison between groups. Thus, these analyses reveal regions, on a subject-by-subject basis, in which CS Valence-dependent BOLD signal change during the PIT stage was associated with aversive PIT. This association was assessed as a function of Action Context and Group. These analyses were repeated with PCL-R score (instead of behavioral PIT) as a covariate to assess whether CS Valence dependent BOLD signal change during the PIT stage was associated with psychopathy severity.

Next, additional analyses were performed to assess whether positive behavioral and fMRI findings from the PIT stage could be explained by BOLD signal change in the Pavlovian conditioning stage. Thus, we assessed in which regions individual differences in CS Valence-dependent BOLD signal change during the Pavlovian training were associated with individual differences in PCL-R score, aversive PIT and neural signaling in the caudate nucleus (see results) during the PIT stage.

First, at the subject level, a GLM was specified with six main regressors of interest representing the onset of the CS trials (during which no US was presented) in the beginning and the end of the conditioning stage: Valence (appetitive/neutral/aversive) X TIME (early/late). This latter distinction between early and late acquisition, was based on evidence of rapid habituation of the responses in the amygdala during conditioning (Birbaumer et al., 2005). Early trials were the first three trials following the first US presentation for aversive and appetitive CS trials. For the neutral CS, the early trials were the first three presentations and the late trials were all the remaining CS presentations thereafter. To capture the other parts of the Pavlovian training, four regressors were added: for appetitive and aversive US onsets, for

juice delivery onset (duration 2 sec) and for the query trial onset (duration 2 sec). Realignment and high-pass filtering was applied as before. Parameter estimates were obtained by maximum-likelihood estimation (AR1). We calculated the main effect of CS Valence (i.e. neutral-aversive) at the subject level for early, late and overall [early+late] conditioning and correlated the effects at the group-level (one sample t-test with covariate) with the PCL-R score; behavioral aversive PIT; and the extracted betas of the caudate nucleus (Figure 5.3) as covariates of interest.

Regions of interest analysis

We report those effects that survive family wise error (FWE) correction for multiple comparisons across the whole brain (P_{WB} <.05, voxel-level) or regions of interest (ROIs): The amygdala(Tzourio-Mazoyer et al., 2002) and nucleus accumbens(Geurts et al., 2013a) were chosen as ROIs based on their known key role in (aversive) PIT(Corbit, 2005; Talmi et al., 2008; Corbit and Balleine, 2011; Prevost et al., 2012; Geurts et al., 2013a). The bilateral caudate nucleus, bilateral putamen and ventromedial prefrontal cortex were chosen based on their role in controlling instrumental action (see introduction). These same regions were also used in Geurts et al.(Geurts et al., 2013a). Following this prior work, we used the action-specific (approach>withdrawal) activation cluster (p<.001 uncorrected; peak voxel MNI-coordinates: [-8,36,-8]) revealed by this previous PIT study(Geurts et al., 2013a) to assess action specific signal in the vmPFC. The left and right elements of the bilateral volumes of interest were combined using Marsbar™(Brett et al., 2002).

5.3 RESULTS

5.3.1 *Behavioural data*

Instrumental and Pavlovian stage

Healthy controls had a tendency to learn faster than the psychopathic criminals (PCs) during instrumental training (Group x Time $F_{1;31}$=4.3, p=.072), but they were matched in terms of instrumental performance during the PIT stage (main effect of Group: $F_{1;31}$=2.0, P =.17). There were no other relevant group differences in these first two stages (Supplementary Materials).

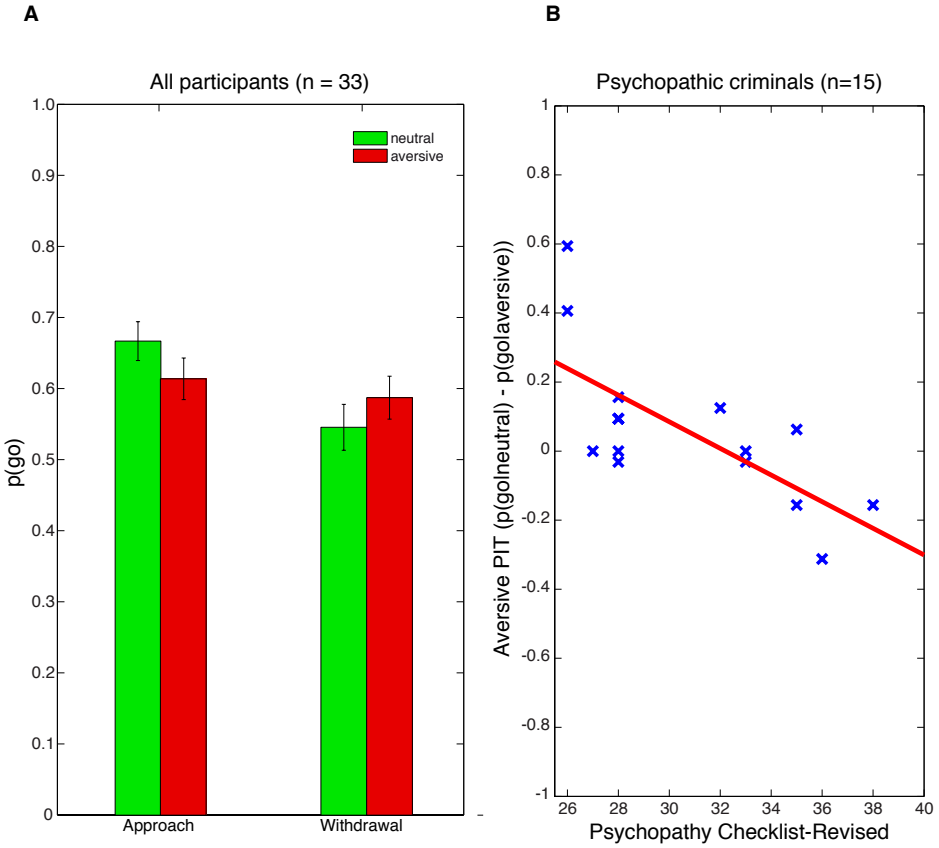


Figure 5.2 *– Behavioural data from the Pavlovian-instrumental transfer stage. Shown are (A) choice (p(go)) as a function of Action Context (approach and withdrawal) and Valence (neutral/aversive) for all participants. Error bars represent standard errors of the means. (B) The correlation between aversive PIT (i.e. p(go | neutral)-p(go | aversive)) and psychopathy severity (in terms of the psychopathy checklist – revised total score). The red line is the ordinary least square trend line. Each cross represents an individual data point.*

Pavlovian-instrumental transfer

There was action-specific PIT (Huys et al., 2011; Geurts et al., 2013a) in that aversive stimuli inhibited approach, but promoted withdrawal (Figure 5.2A; interaction Action Context(approach/withdrawal) x Valence(neutral/aversive): $F_{1;31}$=8.6, p=.006; Valence during approach: $F_{1;31}$=4.7, p=.037; Valence during withdrawal: $F_{1;31}$=4.3, p=.046). Subjects tended to make more go-responses in the approach than in the withdrawal context overall (main Action Context effect: $F_{1;31}$=9.9, p =.004). No significant main effect of or interaction with Group were found (F<2.4, p>.161,Supplementary Table 5.3).

Higher psychopathy severity scores were associated with less aversive PIT (interaction PCL-R score x Valence(neutral/aversive): $F_{1;13}$=12.6, p=.004; Figure 5.2 B). This appeared to be driven by a combination of aversive disinhibition in the approach context and enhanced activation in withdrawal context (Supplementary Results).

5.3.2 *Imaging data*

The neural correlates of aversive PIT did not differ between groups (no significant Action Context(approach/withdrawal) x Valence(neutral/aversive) x Group(PCs/HCs) interactions). Action-specific signals(approach vs. withdrawal) across CS Valence were found (precuneus [12,-78,6], k=5453, Z=7.02, p_{FWE}<.001; lingual [10,-52,52], k=310, Z=5.50, pFWE=.001; and middle occipital gyrus ([34,-88,2], k=159, Z=4.98, p_{FWE}=.016, whole brain corrected).

There was a significant brain-behaviour association: Individual differences in behavioural aversive PIT (p(go|neutral)-p(go|aversive)) correlated positively with BOLD signal in the striatum (aversive main regressor–neutral main regressor). Subjects showing reduced behavioural aversive inhibition also exhibited less signal in the left putamen during neutral versus aversive trials (Figure 5.3A, peakvoxel MNI-coordinates [-26,2,6], k=205, Z=4.59, p_{FWE}=.004, small volume corrected), so that greater suppression of behaviour was accompanied by greater suppression of putamen signal. This effect was present across both groups. Conversely, there was a significant difference between groups in the caudate nucleus (peakvoxel MNI-coordinates [14,20,10], k=98, Z=4.26, p_{FWE}=.006, small volume corrected, Figure 5.3B): Signal in the caudate nucleus (neutral-aversive) also correlated positively with aversive PIT in the PCs (peakvoxel MNI-coordinates [14,18,10], k=225, Z=4.85, p_{FWE} =.039, corrected for the whole-brain). By contrast, in healthy control subjects this brain-behaviour association was not significant and, if anything, in the opposite direction ($p_{uncorrected}$=.012

at peak voxel [14,18,10], Figure 5.3B). In addition, there was no significant brain-PCL-R correlation at the whole brain or in the small volume of the caudate nucleus ($p_{uncorrected}$=.006, $p_{corrected}$>.05).

Next, we asked whether the observed, psychopathy severity related PIT effects (Figure 5.2B and Figure 5.3) were driven by differences in Pavlovian conditioning. PCL-R score was indeed related to CS-dependent BOLD signal change (neutral vs. aversive over the whole conditioning stage) in the bilateral amygdala (right amygdala: peakvoxel MNI-coordinates [24 2 -24], k=22, Z=3.51, p_{FWE}=.033, left amygdala: peakvoxel MNI-coordinates [-18 -2 -22], k=37, Z=3.48, p=.036, small volume corrected for the bilateral amygdala; Figure 5.4). However, CS-dependent amygdala signal during Pavlovian conditioning did not correlate with PIT-related caudate nucleus signal in the PCs ($p_{uncorrected}$>.001), or with aversive PIT behavior ($p_{uncorrected}$>.001). This latter is remarkable because PCL-R scores were strongly associated both with CS dependent amygdala BOLD response during Pavlovian conditioning (Figure 5.4) and with behavioral aversive PIT (Figure 5.2) separately.

5.4 DISCUSSION

The present study shows that within a group of psychopathic criminals higher levels of psychopathy were accompanied by attenuated inhibition of instrumental behavior by aversive cues. Thus, criminals with higher psychopathy scores exhibited reduced aversive PIT. Using fMRI, we demonstrate an aberrant positive association between aversive PIT and neural signaling in the caudate nucleus in the PCs. In this group, aversive inhibition of action was accompanied by aversive inhibition of signaling in the caudate nucleus. Intriguingly, this was observed exclusively in the group of PCs and did not extend to the healthy controls. These data support the hypothesis that psychopathy severity is associated with reduced aversive inhibition of instrumental behavior. The neural data raise the hypothesis that this anomaly reflects aberrant transfer of aversive value onto the caudate nucleus. Together these data establish a key link between psychopathic severity, aversive inhibition and the striatum.

We explicitly focused on the transfer of aversive value to instrumental action, rather than on aversive processing per se. A focus on this transfer of aversive value onto instrumental action is directly relevant to understanding the behavioural anomalies in psychopathy, such as instrumental aggression We argue that it is unlikely that the behavioural transfer effects related to psychopathy severity were due to abnormal aversive processing or Pavlovian conditioning in itself. The transfer effects were not accompa-

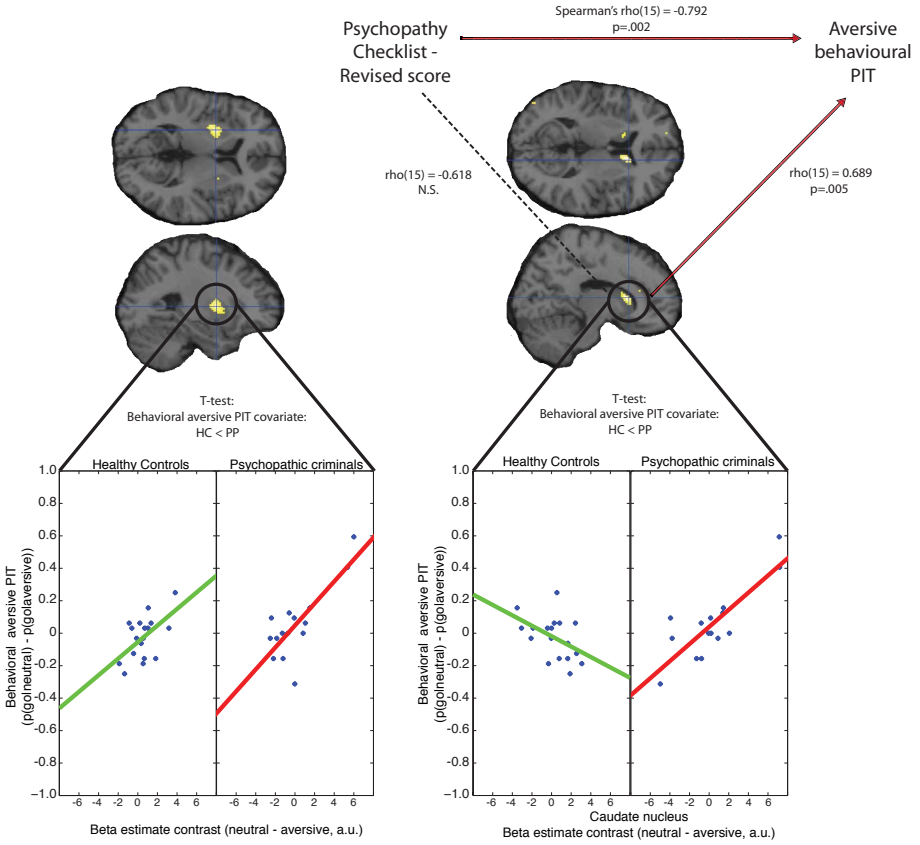


Figure 5.3 *– Association between aversive behavioural PIT, beta estimate contrasts (neutral – aversive trials) and psychopathy severity (in terms of the psychopathy checklist – revised total score) for the putamen and the caudate nucleus.* ***Left panel****(A) Beta estimate contrasts (neutral – aversive trials) within the putamen are positively correlated with behavioural aversive PIT (p(go | neutral) – p(go | aversive)) for psychopathic criminals and healthy controls.* ***Right panel****(B) Beta estimate contrasts (neutral-aversive trials) within the caudate nucleus correlate positively with behavioural aversive PIT (p(go | neutral – p(go | aversive)) for psychopathic criminals, but not for healthy controls. Correlations between the mean beta estimate contrasts and PCL-R and aversive behavioural PIT are calculated in terms of Spearman's rho. Scatterplots are for illustrative purposes only and were created by plotting the behavioural PIT effect against the extracted average beta estimate contrast from the p < .001 whole-brain uncorrected clusters within the putamen and caudate nucleus.*

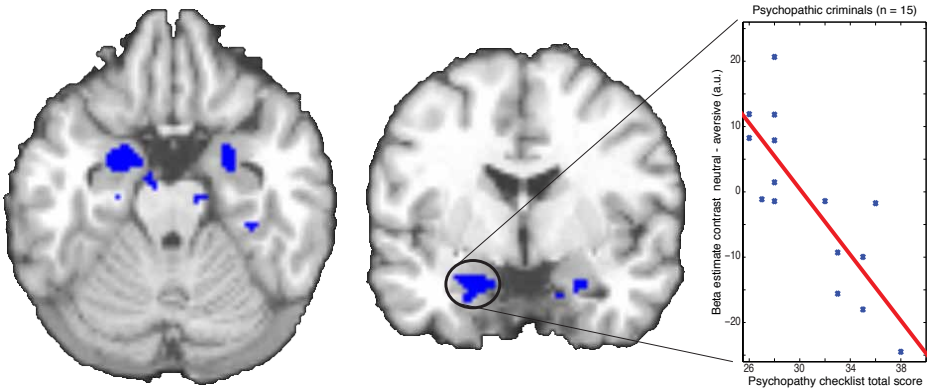


Figure 5.4 *– Association between psychopathy severity (in terms of PCL-R total score) and beta estimate contrasts (CS neutral – CS aversive trials) during Pavlovian conditioning. Scatterplot depicts relation between PCL-R total score and the extracted average beta estimate contrast from the $p < .001$ whole-brain uncorrected cluster within the amygdala (for illustrative purpose only).*

nied by changes in Pavlovian conditioning, as indexed by the query trials, the subjective (liking) ratings and amygdala BOLD signal: The amygdala of criminals with higher psychopathy severity scores was in fact more reactive to aversive CSs compared with neutral CSs than those with lower PCL-R scores (Figure 5.4). This does not support the idea that PCs with higher psychopathy severity scores are more insensitive to aversive CSs and it is in line with recent findings that show increased amygdala BOLD signal in psychopathy (Schultz et al., 2016). Furthermore, increased psychopathy severity predicted reduced aversive PIT in both the approach and withdrawal contexts, with people with high psychopathy scores exhibiting more withdrawal as well as more approach actions. Impairment in aversive Pavlovian conditioning would have surfaced as reduced inhibition of approach, but also as reduced potentiation of withdrawal actions. On the contrary, the potentiating effect of the aversive CS on withdrawal was more rather than less pronounced in subjects with higher PCL-R scores. Note in addition, that there was no significant association between effects of aversive Pavlovian conditioning in the amygdala on the one hand and behavioral aversive PIT and PIT-related signal in the caudate nucleus on the other hand. This is remarkable, because amygdala BOLD during aversive Pavlovian conditioning and behavioral aversive PIT were both strongly related to psychopathy severity. Furthermore, the behavioural transfer effects were also not accom-

panied by any effects in the amygdala during the PIT stage, but rather by aberrant recruitment of the striatum. Altogether, this strengthens the hypothesis that the diminished aversive inhibition found in subjects with higher psychopathy severity scores does not reflect an abnormality in aversive processing systems, such as the amygdala(cf. Blair, 2005), but instead might reflect abnormal transfer of such processing to striatal systems that regulate instrumental control.

We found an anomalous positive association between aversive inhibition of action and aversive inhibition of BOLD signal in the caudate nucleus in PCs. In controls, greater aversive inhibition of behaviour was associated with greater inhibition of BOLD signal only in the putamen. Conversely, inhibition of behaviour in PCs involved inhibition of signal also in the caudate nucleus. In line with evidence from work with experimental rodents, recent evidence in humans(Balleine and Doherty, 2009) implicates the caudate nucleus (together with the ventromedial prefrontal cortex(Valentin et al., 2007)) in the goal-directed control of behaviour(Tanaka et al., 2008; de Wit et al., 2012). This goal-directed form of control is more flexible, but computationally more costly than the more rigid habitual form of behavioural control, which is thought to implicate rather the putamen(Tricomi et al., 2009; de Wit et al., 2012). Work with rodents has suggested that PIT is stronger when behaviour is under habitual than goal-directed control(Holland, 2004). Accordingly, one might speculate, based on our neural findings, that aversive Pavlovian disinhibition occurs in psychopathy due to excessive recruitment of the goal-directed control system, which might be less sensitive to Pavlovian biasing (but see Allman et al., 2010). This possibility requires testing in future studies, for example by combining current (neurocomputational) methods for assessing the relative contributions of these two control systems(Daw, 2011) with a Pavlovian cue manipulation and the study of psychopathy (cf. Sebold et al., 2016).

The finding that increased psychopathic severity was associated with reduced aversive PIT is remarkably in line with the findings of our previous study, which demonstrated, using a comparable experimental task, reduced aversive PIT after central serotonin depletion in healthy volunteers(Geurts et al., 2013b). Criminal psychopathy has been shown to be accompanied by reduced central serotonin transmission, as indexed by reduced serotonin metabolites in the cerebrospinal fluid(Soderstrom et al., 2001; 2003). In line with this observation, we found a strong correlation between the PCL-R-score and aversive PIT. The next step will be to assess whether aversive Pavlovian disinhibition in psychopathy can be countered by serotoninergic drugs, such as selective serotonin reuptake inhibitors, consistent with recent

findings that provoked aggression in primary psychopathy can be reduced by serotonin augmentation by paroxetine (Fanning et al., 2014).

We highlight the following limitations of the current study: First we were puzzled to find that the continuous associations with psychopathy severity rating in terms of PCL-R total score were not accompanied by group differences. The range of behavioural PIT scores in the PCs was comparable with that in the controls. We did not obtain PCL-R-scores from the controls and therefore cannot exclude that a similar association exists in healthy controls. However, we think this is unlikely, because there were no correlations between scores on the Psychopathic Personality Inventory and aversive PIT (Supplementary Results). An alternative possibility is that abnormal aversive PIT per se is not a sufficient prerequisite for developing criminal psychopathy. For example, criminal psychopathy might surface only if abnormal aversive PIT is accompanied by excessive impact of reward on behaviour and cognition (Buckholtz et al., 2010; Bjork et al., 2012; Yildirim and Derksen, 2015; Geurts et al., 2016).

Second, we were not able to replicate the strong action-specific BOLD signal found in the ventromedial prefrontal cortex as observed in our previous fMRI study with the same paradigm in healthy young volunteers(Geurts et al., 2013a). We have recently replicated this ventromedial effect in young women (both healthy and with borderline personality disorder) in another independent dataset (unpublished findings, D.E.M. Geurts, T.J. van den Heuvel, R. Geurts). One factor that might account for this discrepancy is that the average performance at the end of the instrumental task was significantly better in the latter studies (with mainly graduate students) compared with that of the healthy controls in the current study (mean accuracy (SEM): 2013 study =.76 (.023), current = .64 (.027), t-test: t_{36}=3.6, p=.001). Thus it is possible that the healthy controls and patients in the current study relied to a lesser degree on a goal-directed control strategy and to a greater degree on a habitual control strategy than did the subjects in our previous study. Although speculative, this might explain why the putamen was recruited as a function of aversive PIT in the current study in both the healthy control group and the psychopathy group, which was not the case in our previous study. Relevant in this context might also be the fact that the current study included only men, whereas the other studies mainly included women.

Third, we should note that our group comparison is necessarily confounded by overt criminal history. As such, we cannot and do not claim specificity of our findings to psychopathic criminals compared with non-psychopathic criminals or psychopathic non-criminals (cf. "successful psychopaths").

Finally, the insensitivity of our paradigm to appetitive PIT (cf. Geurts et al. (2013a), see supplementary Results) precludes conclusions about the valence-specificity of the effects. This should be addressed in future studies.

5.5 REFERENCES

ALLMAN MJ, DeLeon IG, Cataldo MF, Holland PC, Johnson AW (2010) Learning processes affecting human decision making: An assessment of reinforcer-selective Pavlovian-to-instrumental transfer following reinforcer devaluation. Journal of Experimental Psychology: Animal Behavior Processes 36:402–408.

BALLEINE BW, Doherty JPOA (2009) Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. Neuropsychopharmacology 35:48–69.

BIRBAUMER N, Veit R, Lotze M, Erb M, Hermann C, Grodd W, Flor H (2005) Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. Arch Gen Psychiatry 62:799–805.

BJORK JM, Chen G, Hommer DW (2012) Psychopathic tendencies and mesolimbic recruitment by cues for instrumental and passively obtained rewards. Biological Psychology 89:408–415.

BLAIR RJ (2001) Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. Journal of Neurology, Neurosurgery, and Psychiatry 71:727–731.

BLAIR RJR (2005) Applying a cognitive neuroscience perspective to the disorder of psychopathy. :1–27.

BLAIR RJR (2008) The amygdala and ventromedial prefrontal cortex: functional contributions and dysfunction in psychopathy. Philosophical Transactions of the Royal Society B: Biological Sciences 363:2557–2565.

BRAINARD DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

BRETT M, Anton J-L, Valabregue R, Poline J-B (2002) Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain.

BROOK M, Brieman CL, Kosson DS (2013) Clinical Psychology Review. Clinical Psychology Review 33:979–995.

BUCKHOLTZ JW, Treadway MT, Cowan RL, Woodward ND, Benning SD, Li R, Ansari MS, Baldwin RM, Schwartzman AN, Shelby ES, Smith CE, Cole D, Kessler RM, Zald DH (2010) Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. Nature Neuroscience 13:419–421.

BÜCHEL C, Wise RJ, Mummery CJ, Poline JB, Friston KJ (1996) Nonlinear regression in parametric activation studies. NeuroImage 4:60–66.

CARDINAL RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience and Biobehavioral Reviews 26:321–352.

COCCARO EF (1992) Impulsive aggression and central serotonergic system function in humans: an example of a dimensional brain-behavior relationship. Int Clin Psychopharmacol 7:3–12.

CORBIT LH (2005) Double Dissociation of Basolateral and Central Amygdala Lesions on the General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer. Journal of Neuroscience 25:962–970.

CORBIT LH, Balleine BW (2011) The General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer Are Differentially Mediated by the Nucleus Accumbens Core and Shell. Journal of Neuroscience 31:11786–11794.

DAMASIO AR (1997) Neuropsychology. Towards a neuropathology of emotion and mood. Nature 386:769–770.

DAW N (2011) Model-based Reinforcement Learning as Cognitive Search: Neurocomputational Theories, Unpublished manuscript prepared for the Ernst Strüngmann Forum: Cognitive Search: Evolution, Algorithms, and the Brain February 20–25, 2011

DAYAN P, Niv Y, Seymour B, D Daw N (2006) The misbehavior of value and the discipline of the will. Neural Networks 19:1153–1160.

DAYAN P, Seymour B (2008) Values and actions in aversion. In P. W. Glimcher (Ed.), Neuroeconomics: Decision making and the brain. Academic Press.

DEWIT S, Watson P, Harsay HA, Cohen MX, van de Vijver I, Ridderinkhof KR (2012) Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. Journal of Neuroscience 32:12066–12075.

DOLAN RJ, Dayan P (2013) Goals and Habits in the Brain. Neuron 80:312–325.

ESTES WK, Skinner BF (1941) Some quantitative properties of anxiety. J Exp Psychol 29:390.

FANNING JR, Berman ME, Guillot CR, Marsic A, McCloskey MS (2014) Serotonin (5-HT) augmentation reduces provoked aggression associated with primary psychopathy traits. J Pers Disord 28:449–461.

FLOR H, Birbaumer N, Hermann C, Ziegler S, Patrick CJ (2002) Aversive Pavlovian conditioning in psychopaths: peripheral and central correlates. Psychophysiol 39:505–518.

GEURTS DEM, Borries von K, Volman I, Bulten BH, Cools R, Verkes RJ (2016) Neural connectivity during reward expectation dissociates psychopathic criminals from non-criminal individuals with high impulsive/antisocial psychopathic traits. Social Cognitive and Affective Neuroscience, 11(8), 1326–1334.

GEURTS DEM, Huys QJM, Ouden den HEM, Cools R (2013a) Aversive Pavlovian Control of Instrumental Behavior in Humans. Journal of Cognitive Neuroscience 25:1428–1441.

GEURTS DEM, Huys QJM, Ouden den HEM, Cools R (2013b) Serotonin and Aversive Pavlovian Control of Instrumental Behavior in Humans. Journal of Neuroscience 33:18932–18939.

GLENN AL, Raine A (2009) Psychopathy and instrumental aggression: Evolutionary, neurobiological, and legal perspectives. Int J Law Psychiatry 32:253–258.

HARE RD (2003) Manual for the Hare Psychopathy Checklist-Revised, 2nd ed. Toronto, Ontario, Canada: Multi Health Systems.

HEINZ A, Schlagenhauf F, Beck A, Wackerhagen C (2016) Dimensional psychiatry: mental disorders as dysfunctions of basic learning mechanisms. Journal of Neural Transmission, 123(8), 809–821.

HOLLAND P (2004) Amygdala–frontal interactions and reward expectancy. Current Opinion in Neurobiology 14:148–155.

HUYS QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P (2011) Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding Rangel A, ed. PLoS Comput Biol 7:e1002028.

JELICIC M, Merckelbach H, Timmermans M, Candel I (2004) De Nederlandstalige versie van de Psychopathic Personality Inventory. Psychodiagnostisch gereedschap. De Psycholoog:604–608.

KAHNEMAN D, Frederick S (2007) Frames and brains: elicitation and control of response tendencies. Trends in Cognitive Sciences 11:45–46.

LARSON CL, Baskin-Sommers AR, Stout DM, Balderston NL, Curtin JJ, Schultz DH, Kiehl KA, Newman JP (2013) The interplay of attention and emotion: top-down attention modulates amygdala activation in psychopathy. Cogn Affect Behav Neurosci 13:757–770.

LILIENFELD SO, Andrews BP (1996) Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. J Pers Assess 66:488–524.

LY V, Cools R, Roelofs K (2014) Aversive disinhibition of behavior and striatal signaling in social avoidance. Social Cognitive and Affective Neuroscience 9:1530–1536.

MOUL C, Killcross S, Dadds MR (2012) A model of differential amygdala activation in psychopathy. Psychol Rev 119:789–806.

PORTER S, Woodworth M (2006) Psychopathy and Aggression. In: Handbook of psychopathy, 1st ed. (Patrick CJ, ed), pp 481–494. New York: The Guilford Press.

POSER BA, Versluis MJ, Hoogduin JM, Norris DG (2006) BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. Magn Reson Med 55:1227–1235.

PREVOST C, Liljeholm M, Tyszka JM, O'Doherty JP (2012) Neural Correlates of Specific and General Pavlovian-to-Instrumental Transfer within Human Amygdalar Subregions: A High-Resolution fMRI Study. Journal of Neuroscience 32:8383–8390.

ROTHEMUND Y, Ziegler S, Hermann C, Gruesser SM, Foell J, Patrick CJ, Flor H (2012) Biological Psychology. Biological Psychology 90:50–59.

SCHMAND B, Bakker D, Saan R, Louman J (1991) [The Dutch Reading Test for Adults: a measure of premorbid intelligence level]. Tijdschrift voor Gerontologie en Geriatrie 22:15–19.

SCHULTZ DH, Balderston NL, Baskin-Sommers AR, Larson CL, Helmstetter FJ (2016) Psychopaths Show Enhanced Amygdala Activation during Fear Conditioning. Front Psychol 7:903.

SEBOLD M, Schad DJ, Nebe S, Garbusow M (2016) Don't Think, Just Feel the Music: Individuals with Strong Pavlovian-to-instrumental Transfer Effects Rely Less on Model-based Reinforcement Learning. Journal of Cognitive Neuroscience, 28(7), 985–995.

SERIN RC, Amos NL (1995) The role of psychopathy in the assessment of dangerousness. Int J Law Psychiatry 18:231–238.

SEYMOUR B, Dolan R (2008) Emotion, Decision Making, and the Amygdala. Neuron 58:662–671.

SODERSTROM H, Blennow K, Manhem A, Forsman A (2001) CSF studies in violent offenders I. 5-HIAA as a negative and HVA as a positive predictor of psychopathy. J Neural Transm 108:869–878.

SODERSTROM H, Blennow K, Sjodin AK, Forsman A (2003) New evidence for an association between the CSF HVA: 5-HIAA ratio and psychopathic traits. Journal of Neurology, Neurosurgery, and Psychiatry 74:918–921.

TALMI D, Seymour B, Dayan P, Dolan RJ (2008) Human Pavlovian Instrumental Transfer. Journal of Neuroscience 28:360–368.

TANAKA SC, Balleine BW, O'Doherty JP (2008) Calculating Consequences: Brain Systems That Encode the Causal Effects of Actions. Journal of Neuroscience 28:6750–6755.

TRICOMI E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. European Journal of Neuroscience 29:2225–2232.

TZOURIO-MAZOYER N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. NeuroImage 15:273–289.

VALENTIN VV, Dickinson A, O'Doherty JP (2007) Determining the Neural Substrates of Goal-Directed Learning in the Human Brain. Journal of Neuroscience 27:4019–4026.

VANVLIET IM, Leroy H, van Megen HJGM (n.d.) M.I.N.I Internationaal Neuropsychiatrisch Interview, 5 ed.

VEIT R, Flor H, Erb M, Hermann C, Lotze M, Grodd W, Birbaumer N (2002) Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. Neurosci Lett 328:233–236.

WEERTMAN A, Arntz A, Kerkhofs M (2000) Handleiding gestructureerd klinisch interview voor DSM-IV AS-II persoonlijkheidsstoornissen. Lisse, The Netherlands: Swets test publishers.

WUNDERLICH K, Rangel A, O'Doherty JP (2009) Neural computations underlying action-based decision making in the human brain. Proceedings of the National Academy of Sciences 106:17199–17204.

YILDIRIM BO, Derksen J (2015) Mesocorticolimbic dopamine functioning in primary psychopathy]; A source of within-group heterogeneity. Psychiatry Research, 229(3), 633–677.

YILDIRIM BO, Derksen JJL (2013) Systematic review, structural analysis, and new theoretical perspectives on the role of serotonin and associated genes in the etiology of psychopathy and sociopathy. Neuroscience and Biobehavioral Reviews, 37(7), 1254–1296.

5.6 SUPPLEMENT

5.6.1 *Suplementary Material and Methods*

Forensic psychiatric hospital

The Pompestichting is a "TBS-clinic" located in Nijmegen. TBS ("Ter Beschikking Stelling") is a treatment disposal on behalf of the state for people who committed serious criminal offences in connection with having a mental disorder. TBS is not a punishment, but an entrustment act for mentally disordered offenders (diminished responsibility). These court orders are an alternative to either long term imprisonment or confinement in psychiatric hospital, with the goal to strike a balance between security, treatment and protection.

Additional procedural details

Participants received written and oral information about the experiment and signed an informed consent. All participants were invited for a screening session and a scan session with no more than two weeks in between the appointments. During the first appointment, they were screened for psychiatric exclusion criteria by trained psychologists using the Structure Clinical Interview for DSM disorders to exclude axis 2 disorder (SCID-II; Dutch version, Weertman (Weertman et al., 2000)), Mini International Neuropsychiatric Interview to exclude axis 1 disorder (MINI; Dutch version(van Vliet et al., n.d.)) and the Dutch version of the National Adult Reading Test for IQ assessment (NLV, Schmand (1991)). Psychiatric exclusion criteria were recent major depressive disorder, bipolar disorder, schizophrenia, schizoaffective disorder, schizophreniform disorder, delusional and other psychotic disorders, schizoid or schizotypical personality disorder, current alcohol and substance intoxication, first degree relatives with DSM IV axis I schizophrenia or schizophreniform disorder.

Further, participants completed the PPI(Lilienfeld et al. (1996); Dutch version, Jelicic et al.(2004)). They were instructed not to drink more than 3 units/day during in the week preceding the experimental measure; not to use of alcohol within 24 hours of the measurement; not to use cannabis or other illicit drugs within the week before measurement; not to use psychotropic medication other than oxazepam during the 5 days before measurement; not to use oxazepam within 12 hours before measurement; and not to smoke within 1 hour before measurement and no more than five cigarettes on the scan day. Furthermore, they were asked to refrain from

any caffeinated drinks and chocolate on the scan day and to refrain from extensive physical exercise and heavy meals before the scan session. In the scanner, participants wore earplugs with integrated speakers. Foam pads were placed inside the head coil and paper tape was placed over the forehead and the base of the head coil to restrict movement. Before performing the PIT task, participants performed an approach avoidance task and monetary incentive delay task reported elsewhere. After a break of 15 minutes, they were seated in front of a laptop and they clicked through the same instructions they would receive within the scanner. The investigator who sat next to the participants during these instructions answered possible questions. Instructions and task images were then projected onto a translucent screen at the end of the scan tube, which was visible via a mirror attached to the head coil. After again displaying the instructions of the task, the task was started, which lasted about 50 minutes.

Additional task details

The paradigm was programmed using Matlab (2009b, TheMathWorks, Natick, MA) with the Psychophysical Toolbox extension(Brainard, 1997).

Additional information on the instrumental stage

To orthogonalize the approach-withdrawal and appetitive-aversive axes, the learned instrumental values in approach and withdrawal needed to be matched. To achieve this, both go and no-go responses were, if correct, rewarded to the same extent. Additionally, to avoid confound of behavioural activation, in each condition (i.e., in both approach and withdrawal conditions) the go response was designated as the correct response for half of the instrumental stimuli, and the no-go response for the other half. Incorrect responses had opposite outcome contingencies to correct responses, yielding more punishments than rewards. This ensured that go, no-go, approach, and withdrawal overall had the same learned association with rewards and punishments. In both the approach and withdrawal action context, there were two go stimuli, which yielded reward more often after active responses (and punishment after not responding), and two no-go stimuli, which yielded reward more often after not responding (and punishment after go responding) (on average the ratio reward:punishment after a correct action was 0.86:0.14 for go-stimuli and 0.84:0.16 for nogo-stimuli). Trials were labelled as correct if subjects chose the usually rewarded response. Average reinforcement was matched between approach and withdrawal contexts (mean proportion of positively reinforced trials for approach = 0.59; for withdrawal =0.57, paired sample T-test: T_{33}=1.3, p=.2). Accordingly, the instrumental stimuli in the approach and withdrawal conditions did not

differ in acquired value, so that any differences between conditions cannot reflect differential Pavlovian responses elicited by these instrumental stimuli. Rather than representing effects of competing Pavlovian responses, the effects we report represent PIT effects, i.e. the effects of Pavlovian CSs on instrumental behaviour in terms of choice (percentage of go-choices) and vigour (average number of button presses on go-trials). Initial stimuli and action context were randomized across participants.

Image Acquisition

Whole-brain imaging was performed on a 3-Tesla MR scanner (Magnetrom Trio Tim, Siemens Medical Systems, Erlangen, Germany). Functional data were obtained using a multiecho gradient T2*-weighted EPI (ME-EPI) scanning sequence(Porter and Woodworth, 2006) with BOLD contrast (38 axial-oblique slices; repetition time = 2.32 sec; echo times = 9.0, 19.3, 30, and 40 msec; in plane resolution = 3.3 × 3.3 mm; slice thick- ness = 2.5 mm; distance factor = 0.17; flip angle = 90°, 194 volumes per run acquired during the PIT stage). Visual stimuli were projected on a screen and were viewed through a mirror attached to the head coil. In addition, a high-resolution T1-weighted magnetization- prepared rapid-acquisition gradient-echo anatomical scan was obtained from each participant (192 sagittal slices; repetition time = 2.3 sec; echo time = 3.03 msec; voxel size = 1.0 × 1.0 × 1.0 mm; field of view = 256 mm).

5.6.2 *Supplementary Results*

Instrumental stage

Subjects learned to make correct choices during the instrumental learning stage indicated by an increasing number of correct responses across time (main effect of Time: $F_{1;31}$=27.4.0, p<.001) (Supplementary Figure 5.5A). There was a significant Action Context x Resonse Type interaction ($F_{1;31}$=60.6, p <.001). This was driven by an effect in the approach Action Context where subjects performed better on approach-go stimuli compared with approach-nogo stimuli (F1;31=36.7, p<.001), whereas in the withdrawal Action Context subjects performed better on withdrawal-nogo stimuli compared with withdrawal-go stimuli ($F_{1;31}$=6.6, p=.015). In addition, across Time subjects made more correct go responses than correct nogo responses (main effect of Response Type: $F_{1;31}$=4.5, p=.045).

Performance at the end of instrumental training generalized to, and persisted throughout the PIT stage (Supplementary Figure 5.5): there were no

significant main effects of, or interactions with a Time factor with 3 levels: the end of the instrumental training, the beginning of the PIT stage and the end of the PIT stage ($F_{2,62}<1.5$, $p>.24$). Other patterns found during the instrumental training remained significant during the PIT stage (main effect of Response Type: $F_{1,31}=12.6$, $p=.001$; Action Context x Response Type interaction: $F_{1,31}=13.7$, $P=.001$). There were no significant interactions with or a main effect of the factor Group ($F_{2,62}<2.5$, $p>.12$). Critically, there were also no performance differences between groups when the PIT stage was analysed by itself (main effect of Group: $F_{1,31}=2.0$, $P=.17$).

NO RELATION BETWEEN PERFORMANCE IN THE INSTRUMENTAL STAGE AND AVERSIVE PIT In order to analyse whether performance (i.e. accuracy, p(correct)) during the instrumental stage did influence the outcome of the PIT stage, 1) the performance at the end of the instrumental training and 2) the difference in performance between the beginning and the end of the instrumental training were added as covariates to the rmANOVA (Group x Action Context xValence). If anything, adding these covariates increased the significance of the action specific aversive PIT effect across both groups and neither resulted in any interactions with performance. In addition, this analysis revealed that across groups, subjects who performed better at the end of instrumental training showed less go-actions in the PIT stage ($F_{1,30}=6.7$, $p=.015$).

Pavlovian stage

Data for the VAS-rating analysis were not available for 1 healthy control subject (had to leave earlier) and 3 PCs (one who only finished the first block and two due to technical error at the second post-conditioning rating). Three PCs were excluded from this analysis, because they failed to answer more than half of the query trials in time (2 sec). Results from the analysis of our primary (behavioural and imaging) effects of interest did not change when we excluded the subjects who did not complete the query trials. Furthermore, there was no significant correlation between PCL-R score on the one hand and performance on the query trials or number of missed query trials on the other hand (Spearman's rho13 = .54, p=.19). There was also no performance difference between groups on the Pavlovian query trials, suggesting that explicit CS-US associations were unaffected (Mann-Whitney U test: mean proportion correct across blocks in PCs (n=13): 85%; SEM: 4.9; range: 50-100%; HCs(n=18): 88%; SEM: 3.4; range: 63-100%, p=1.0).

VAS (liking) ratings for the Pavlovian CSs showed that the Pavlovian conditioning procedure induced changes in subjective liking and that there

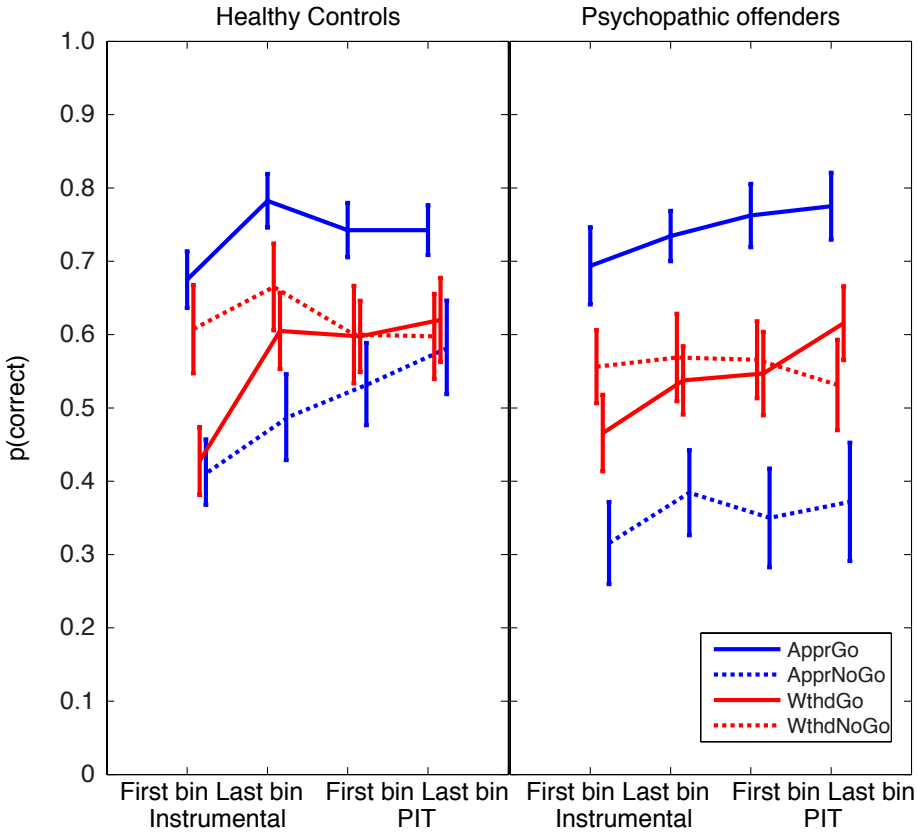


Figure 5.5 *– Instrumental learning and generalization to the Pavlovian-instrumental transfer stage for healthy controls (left panel) and PCs (right panel). The proportion of correct choices (p(correct)) are broken down by RESPONSE TYPE (go/nogo) and ACTION context (approach/withdrawal). Error bars represent standard errors of the mean.*

were no differences between the groups: The aversive CS became more aversive (Wilcoxon Signed Rank Test: p<.001) and the neutral CS did not change (Wilcoxon Signed Rank Test: p=.22). After conditioning the aversive CS was judged to be more aversive than the neutral (Mann-Whitney U test: p=.001). None of the VAS ratings for the Pavlovian CSs or their changes from before to after conditioning differed between the groups (All Mann-Whithney U tests: p>.19). These VAS ratings suggest robust Pavlovian conditioning that does not significantly differ between groups. None of the (changes in) VAS ratings were significantly related to psychopathy severity.

Mann-Whitney U tests showed that there were no differences between the groups in how they rated the aversive juice before and after conditioning (for all comparisons p>.61).

MAIN RESULTS WERE ROBUST TO EXCLUSION OF SUBJECTS WITH ABERRANT PERFORMANCE ON QUERY TRIALS After excluding three subjects who failed to respond to most of the query trials the correlation between PCL-R and aversive Pavlovian inhibition remained strong and significant: Spearman rank correlation: rho_{12}= -.730, p=.007. This held also for the correlation between PCL- R and mean betas from the caudate nucleus: Spearman rank correlation: rho12 = -.67, p=.017; and for the correlation between PCL-R and mean betas form the putamen for the whole group: rho_{30}=.49, p=.006.

Further significant contrasts and their reported correlations from the imaging analysis did not change substantially when excluding these subjects.

Pavlovian to instrumental transfer stage

VIGOUR Analysis of vigour (i.e. number of button presses) revealed a main effect of Valence (neutral vs aversive) (Supplementary table 5.4, $F_{1;31}$=10.8, p=.002). There were no significant differences between the groups in terms of the vigour of responding and the effect of CS Valence on vigour did not depend on individual differences in clinical PCL-R ratings (Valence x PCL-R: $F_{1;13}$=3.0, p=.106).

PSYCHOPATHY PERSONALITY INVENTORY In addition to PCL-R scores (obtained only in PCs) we obtained PPI scores in both groups. However, we did not find any significant association between behavioural aversive PIT and either the PPI total score or the 2 factor model subscores (all p>.1) (across both groups, and in each group separately).

	Healthy controls (n=20)		PCs (n=15)	
	Before	**After**	**Before**	**After**
Appetitive	.52 (.040)	.60 (.026)	.52 (.048)	.61 (.032)
Neutral	.47 (.046)	.52 (.033)	.49 (.056)	.49 (.040)
Aversive	.53 (.039)	.34 (.044)	.48 (.047)	.38 (.053)

Table 5.2 *– Visual analogue scale ratings before and after Pavlovian conditioning (1=very nice, 0=very aversive, mean [SEM]).*

	Healthy controls (n=20)		PCs (n=16)	
	Approach	**Withdrawal**	**Approach**	**Withdrawal**
Appetitive	.652 (.037)	.498 (.042)	.773 (.047)	.551 (.047)
Neutral	.647 (.037)	.525 (.048)	.748 (.042)	.588 (.053)
Aversive	.578 (.043)	.623 (.044)	.676 (.048)	.609 (.050)

Table 5.3 *– Choice (p(go) as a function of Action Context and CS Valence (mean [SEM]).*

BREAKDOWN OF PCL-R AND AVERSIVE PIT RELATION To provide further insight in the Valence x PCL-R relation (Figure 5.2B) we assessed this relation for approach and withdrawal separately. Breakdown of the Valence x PCL-R interaction showed that a higher PCL-R score was accompanied by attenuated inhibition in the approach context in reaction to the aversive compared to the neutral CS ((p(go | neutral&approach)-p(go | aversive&approach)) X PCL-R score: $F_{1;13}$=5.6, p=.035) and in trend by enhanced activation in the withdrawal context ((p(go | neutral&withdrawal)-p(go | aversive&withdrawal)) X PCL-R score: $F_{1;13}$=4.3, p=.059).

APPETITIVE PIT In supplementary analyses, we confirmed that the paradigm was not sensitive to appetitive PIT (neutral vs. appetitive). The rmANOVA with Action Context(approach/withdrawal) and CS Valence(neutral/appetitive) as within subject factors and Group(HCs/PCs) as between subject factor did not reveal any significant PIT-effects not with choice (p(go), all F<1.7, all p>.05) and not with vigour (number of button presses, all F<3.4, all p>.05) as dependent variable.

	Healthy controls (n=20)		PCs (n=16)	
	Approach	**Withdrawal**	**Approach**	**Withdrawal**
Appetitive	7.73 (.44)	7.43 (.53)	8.74 (.49)	7.86 (.59)
Neutral	7.82 (.40)	7.88 (.51)	8.49 (.45)	8.00 (.57)
Aversive	7.23 (.51)	7.78 (.50)	8.09 (.57)	8.00 (.56)

Table 5.4 *– Number of button presses during go-trials as a function of ACTION context and CS VALENCE (mean [SEM]).*

5.6.3 *Supplementary References*

BRAINARD DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

JELICIC M, Merckelbach H, Timmermans M, Candel I (2004) De Nederlandstalige versie van de Psychopathic Personality Inventory. Psychodiagnostisch gereedschap. De Psycholoog:604–608.

LILIENFELD SO, Andrews BP (1996) Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. J Pers Assess 66:488–524.

PORTER S, Woodworth M (2006) Psychopathy and Aggression. In: Handbook of psychopathy, 1st ed. (Patrick CJ, ed), pp 481–494. New York: The Guilford Press.

SCHMAND B, Bakker D, Saan R, Louman J (1991) [The Dutch Reading Test for Adults: a measure of premorbid intelligence level]. Tijdschrift voor Gerontologie en Geriatrie 22:15–19.

VANVLIET IM, Leroy H, van Megen HJGM (n.d.) M.I.N.I Internationaal Neuropsychiatrisch Interview, 5 ed.

WEERTMAN A, Arntz A, Kerkhofs M (2000) Handleiding gestructureerd klinisch interview voor DSM-IV AS-II persoonlijkheidsstoornissen. Lisse, The Netherlands: Swets test publishers.

6

NEURAL CONNECTIVITY DURING REWARD EXPECTATION DISSOCIATES PSYCHOPATHIC CRIMINALS FROM NON-CRIMINAL INDIVIDUALS WITH HIGH IMPULSIVE/ANTISOCIAL PSYCHOPATHIC TRAITS

Geurts, D. E. M., Borries, von, K., Volman, I., Bulten, B. H., Cools, R., & Verkes, R. J.

Social Cognitive and Affective Neuroscience, 2016, 11(8), 1326–1334.

Criminal behaviour poses a big challenge for society. A thorough understanding of neurobiological mechanisms underlying criminality could optimize its prevention and management. Recently, it has been proposed that neural mechanisms underpinning reward expectation might be pivotal to understanding criminal behavior. However this proposal has not been tested in a criminal sample. To fill this gap, we assessed reward expectation in incarcerated, psychopathic criminals. We compared this group to two groups of non-criminal individuals: one with high levels and another with low levels of impulsive/antisocial traits. Functional magnetic resonance imaging (fMRI) was used to quantify neural responses to reward expectancy. Psychophysiological interaction analyses were performed to examine differences in functional connectivity patterns of reward-related regions. The data suggest that overt criminality is characterized, not by abnormal reward expectation per se, but rather by enhanced communication between reward-related striatal regions and frontal brain regions. We establish that incarcerated psychopathic criminals can be dissociated from noncriminal individuals with comparable impulsive/antisocial personality tendencies based on the degree to which reward-related brain regions interact with brain regions that control behavior. The present results help us understand why some people act according to their impulsive/antisocial personality while others are able to behave adaptively despite reward-related urges.

6.1 INTRODUCTION

Criminal behaviour causes great individual suffering as well as large social and economic costs (Wickramasekera et al., 2015). There is a pressing need to understand this behaviour to improve risk assessment, prevention and treatment strategies (van der Gronde et al., 2014). Here we add to this understanding by advancing recent insights in the neurobiology of reward processing derived from studying impulsive/antisocial traits in healthy community samples compared to a criminal sample. In the perspective of risk assessment regarding recurrent criminal behaviour, the construct of psychopathy is of particular interest. It is highly associated with violent criminal behavior (Porter and Woodworth, 2006; Blais et al., 2014): For example, people fulfilling the criteria for psychopathy are overrepresented in the US prison population: about 25% of inmates are diagnosed with psychopathy compared with 1% of the general population (Hare, 2003; Porter and Woodworth, 2006). It might therefore not be surprising, that tools developed to assess psychopathy have found to be useful in predicting future criminal behaviour (e.g. Camp et al., 2013; Whittington et al., 2013). Especially the impulsive/antisocial factor of psychopathy has repeatedly been shown to be predictive of violence (e.g. Edens et al., 2008; Kennealy et al., 2010; Camp et al., 2013; Blais et al., 2014).

Interestingly, recent advances in neurobiological research elucidate the neural underpinnings of this impulsive/antisocial factor (Buckholtz et al., 2010a): Functional MRI and PET evidence suggest that reward expectancy and its underlying mesolimbic dopamine system, might be key to understanding impulsive/antisocial traits (Bjork et al., 2012; Buckholtz et al., 2010a). These seminal findings were however collected from healthy control, community samples and therefore precluded direct conclusions about its relevance for understanding overt criminality. Here we will fill this gap by assessing the neurobiological underpinnings of reward expectancy in a low and high impulsive/antisocial non-criminal group and a (psychopathic) criminal group also scoring high on impulsive/antisocial traits. This will allow us to further our understanding of the relation between impulsive/antisociality, reward expectancy and, critically, overt criminality on a neurobiological level.

More specifically, recent work on reward expectation has shown that non-criminal volunteers with impulsive/antisocial personality traits (assessed with the Psychopathy Personality Inventory (Lilienfeld and Andrews, 1996)) exhibit enhanced reward expectancy-related blood oxygen level dependent (BOLD) signal in the ventral striatum(Buckholtz et al., 2010a; Bjork et al.,

2012) as well as enhanced ventral striatal dopamine release (Buckholtz et al., 2010a). Buckholtz et al. (2010a) used a monetary incentive delay task to assess the association between reward anticipation and impulsive/antisocial traits in a mixed gender community sample. During reward anticipation, the right nucleus accumbens (NAcc) signal correlated positively with the impulsive antisocial factor of psychopathy. The authors proposed that this neural hyper-reactivity to reward expectation is either a direct consequence of aberrant firing of midbrain DA neurons (ventral tegmental area) or a result of decreased regulatory control of ventral striatal activity through a broad inhibitory failure of prefrontal areas. These results have been extended by Bjork et al. (2012), who showed that impulsive/antisocial traits correlate positively not only with ventral striatal activity during instrumentally obtained rewards, but also with anticipation of passively obtained rewards in the anterior mesofrontal cortex. To advance these findings to a forensic level, involving overt and severe criminality, it is pivotal to test criminal, impulsive/antisocial individuals. This enables direct assessment of whether enhanced neural processing of reward expectation in noncriminal impulsive/antisocial adults extends to criminal impulsive/antisocial individuals. Therefore, the aim of this study was to investigate the neural mechanism underlying reward expectation in a group of criminals scoring high on antisocial/impulsivity factor of the psychopathic personality inventory. Specifically, we assessed whether these criminals show similar (or even greater) increases in ventral striatal reward expectancy-related BOLD signal as do (than) non-criminal healthy controls with high impulsive and antisocial traits (following Buckholz et al. 2010a). If the ventral striatal reactivity is related to the level of impulsive/antisociality as measured by the PPI, but not directly related to criminality, we expect no group differences in ventral striatal reactivity between the criminal and non-criminal high impulsive/antisocial groups. In addition, overt criminality might only emerge in high-impulsive antisocial persons if the relatively high level of ventral striatal reactivity to reward expectation is not accompanied by appropriate regulation of other brain areas. We tested this latter hypothesis by assessing differences in neural connectivity between the healthy control group scoring high on impulsive/antisocial traits and the criminal group.

Note that the impulsive/antisocial traits assessed here are an integral part of the psychopathy construct (Neumann et al., 2005; Hare and Neumann, 2008), but do not specifically distinguish psychopathic criminals from other criminals(Patrick et al., 2009). Here we nevertheless focus our analyses on these traits, rather than on the interpersonal and affective traits of psychopathy, firstly because we aim to further the findings of Buckholtz et al.(2010a)

who were able to convincingly couple these traits to reward-expectation and its underlying neurobiology. Second, the aim to advance insight in overt criminality seems to be best served by assessing the impulsive/antisocial traits as measured by the PPI: These traits reflect past violence and predict future violence more consistently and with larger effect sizes than the interpersonal/affective factor (e.g. Edens et al., 2008; Kennealy et al., 2010; Camp et al., 2013; Blais et al., 2014).

6.2 MATERIAL AND METHODS

6.2.1 *Participants*

We assessed BOLD signal with fMRI in 34 subjects using a monetary incentive delay task, known to induce reward-related BOLD signal in the ventral striatum(Knutson et al., 2001; Hoogman et al., 2013). These 34 subjects consisted of 20 healthy subjects without criminal record and 14 psychopathic criminals (Table 6.1). The latter group was part of a group of 18 patients recruited on a voluntary basis from the inpatient population of a high security forensic psychiatric hospital in the Netherlands based on available information about clinical status and prior history. Two criminals had to be excluded due to technical problems and 2 withdrew from participation during the study. The remaining 14 psychopathic criminals were between 18 and 55 years of age (mean age = 40.14, SD = 8.82, 3 left handed) and diagnosed with a psychopathy score of $\geq$ 26 according to the Hare Psychopathy Check List-Revised (PCL-R(Hare, 2003); mean total score = 30.6, SD = 3.9). We assessed IQ levels using the Dutch version of the National Adult Reading Test(Schmand et al., 1991). Twenty healthy men (3 left-handed) matched for age and IQ (mean age = 40.8, SD = 9.86) without criminal records and/or a history of current psychiatric disorders were recruited by advertisement among employees of the high security forensic psychiatric hospital. Participants in both groups were checked for drug use and for medical history. Furthermore, all subjects (n=34) were assessed with the PPI(Lilienfeld and Andrews, 1996). Following Buckholtz et al(2010a), we focused on the second factor (impulsive/antisocial traits) and divided the healthy control group by median split (median = 146) in a high and a low scoring group.

The above procedure resulted in 3 groups: (i) a noncriminal control group with low impulsive/antisocial traits, (ii) a noncriminal control group with high impulsive/antisocial traits and (iii) a psychopathic criminal group (Table 6.1). The 3 groups did not differ in IQ and age (both $p > .715$; Table 6.1). Of note is that impulsive/antisocial trait scores did not differ between the

	PP (n=14)	HClow (n=10) (<146 on PPI_IA)	HChigh (n=10) (>146 on PPI_IA)	Statistics (p-value)
Age	40.1 (8.8)	42.5 (10.22)	39.1 (9.7)	.715
IQ (NLV)	100.1(11.0)	101.30 (10.60)	103.1 (5.4)	.733
PCL-R total	30.6 (3.8)	-	-	-
PCL-R Factor 1	11.9 (2.8)	-	-	-
PCL-R Factor 2	13.9 (2.0)	-	-	-
PPI total	362.2 (42.5)	323.7 (29.8)	368.3 (20.8)	.01
PPI_FD (factor 1)	142.6 (19.9)	133.5 (23.7)	145.6 (14.6)	.366
PPI_IA (factor 2)	165.1 (28.4)	132.8 (9.7)	169.2 (14.7)	.001

Table 6.1 *– Group characteristics (mean, standard deviation) of the group of psychopathic criminals (PP) and healthy matched control subjects scoring high on the PPI_IA factor (HChigh) and healthy matched controls scoring low on the PPI_IA factor (HClow). NLV = Dutch reading test, PCL-R = Psychopathy checklist revised, PP = psychopathy group, PPI_FD = factor 1 of the PPI 'fearless dominance', PPI_IA = factor 2 of the PPI 'impulsive antisociality*

noncriminal healthy high impulsive/antisocial and the criminal psychopathy group (Table 6.1).

6.2.2 *Psychopathy assessment*

In both groups psychopathic traits were assessed using the PPI, and additionally in the criminal group the PCL-R.

Psychopathy Checklist – Revised (PCL-R)

The PCL-R is a 20-item instrument for assessing criminal psychopathy in research, clinical and forensic settings. In contrast to the PPI, this is not a self-report questionnaire, but items are scored by at least 2 independent, trained raters based on file information, collateral reports and extensive interviewing.

Psychopathic Personality Inventory (PPI)

The PPI is a 187 item self-report questionnaire designed to measure psychopathy in community samples(Lilienfeld and Andrews, 1996). Items are answered on a 4-point Likert scale (1 = false, 4 = true). Eight subscales are scored which can be further reduced into 2 factors (respectively fearless-

dominance; and impulsive-antisocial behavior (IA)), which in turn are summed into a total score representing a global index of psychopathy. Note that the IA factor was used to split the healthy control group into high and low scoring subjects.

6.2.3 *Procedure*

Participants received written and oral information about the experiment and signed an informed consent. All participants were invited for a screening session and a scan session with no more than 2 weeks in between the appointments. During the first appointment, they were screened for psychiatric exclusion criteria by trained psychologists using the Structure Clinical Interview for DSM disorders to exclude axis 2 disorder (SCID-II; Dutch version(Weertman et al., 2000)), Mini International Neuropsychiatric Interview to exclude axis 1 disorder (MINI; Dutch version(van Vliet et al., n.d.)) and the Dutch version of the National Adult Reading Test for IQ assessment (NLV(Schmand et al., 1991)). Further, participants completed the PPI(Lilienfeld and Andrews, 1996) (Dutch version(Jelicic et al., 2004)). They were instructed not to drink more than 3 units/day during in the week preceding the experimental measure, not to use of alcohol within 24 hours of the measurement, not to use cannabis or other illicit drugs within the week before measurement, not to use psychotropic medication other than oxazepam during the 5 days before measurement, not to use oxazepam within 12 hours before measurement and not to smoke within 1 hour before measurement and no more than 5 cigarettes on the scan day. Furthermore, they were asked to refrain from any caffeinated drinks and chocolate on the scan day and to refrain from extensive physical exercise and heavy meals before the scan session. In the scanner, participants wore earplugs and headphones. Foam pads were placed inside the head coil to restrict movement and a heartbeat device was connected to the second toe. Before performing the monetary incentive delay task, participants performed an approach avoidance task reported elsewhere. Instructions and task images were projected onto a translucent screen at the end of the scan tube, which was visible via a mirror attached to the head coil. Participants received a practice block that was stopped when participants had 5 on-time responses (hits). After summarizing again the purpose of the task, the experimental block was started, which lasted 12 minutes. After a short break outside the scanner, the anatomical scan (duration: 5 min) and an unrelated task were acquired in the 3T MR scanner, which was located in an adjacent scanner room.

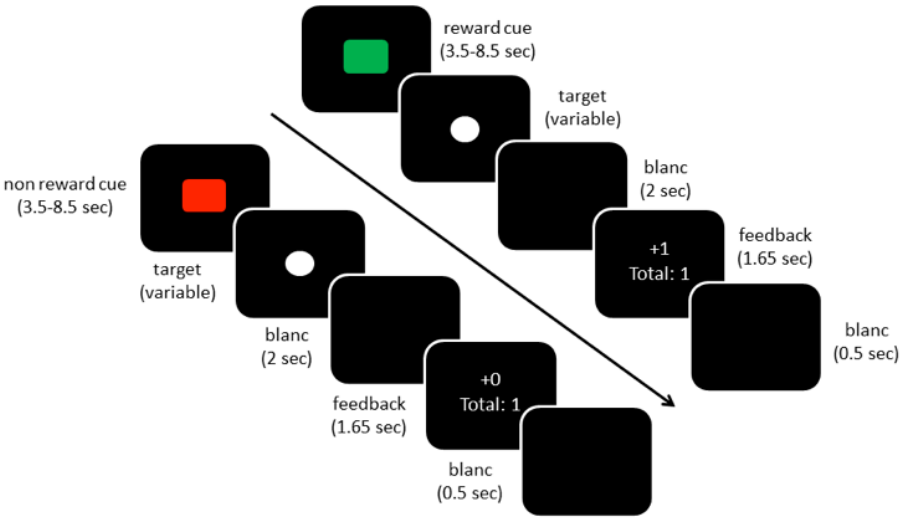


Figure 6.1 *– Task schematics: the upper row showing a reward and the lower row showing a non-reward hit trial (i.e. responses below a variable response limit).*

6.2.4 *Experimental Task*

Monetary incentive delay task

The monetary incentive delay (MID) task (Figure 6.1) consisted of 75 trials (25 potentially rewarding, 25 potentially non-rewarding and 25 baseline fixation trials). Each trial started with a cue (green square indicating reward trials and red square indicating no-reward trials), which was presented for 3500-8500 msec. Next, a white circle was presented (target) to which the participants had to respond as quickly as possible by pressing a button. The target was followed by a black screen for 2000 msec, after which the outcome was displayed (1650 msec) informing the participant about the outcome of the current trial (+/- 1) and the total amount of points. Participants could gain 1 point in the reward condition and no points in the no-reward condition if they responded between 270 and 500 msec after target onset. The response window was adjusted on an individual level and separately for reward and no-reward trials (after a hit, 20 msec was subtracted from the last response window, after a miss, 10 msec was added to the last response window). For every participant, the initial response window was set to 270 msec. This procedure resulted in comparable hit rates in the reward (34%) and no-reward (30%) condition. Before the next cue was shown, a black screen was shown for 500 msec. Participants were told that the total amount of points was converted to monetary rewards (1 point resulted in 20 eurocents) and would be given as a bonus to the regular payment for participation.

6.2.5 *Behavioral analysis*

The number of hits, reaction times on hits and target duration were submitted to a 2 x 3 repeated measures analysis of variance (rmANOVA) with Reward Expectation (reward, no-reward) as within-subject factor and Group (psychopathic criminals, noncriminal high-, noncriminal low- impulsive/antisocial) as between-subject factor.

6.2.6 *MR Image acquisition and analysis*

Image acquisition

Whole-brain imaging was performed on a 1.5 Tesla MR scanner (Magnetrom Sonata, Siemens Medical Solutions, Erlangen, Germany) using an 8-channel coil. Functional data were obtained using a multi-echo gradient

T2*-weighted echo-planar (ME-EPI) scanning sequence(Poser et al., 2006) with BOLD contrast (34 axial-oblique slices, repetition time, 2.64s; echo-times: 6.9, 24.2, 33, 43, and 52 msec; in plane resolution, 3.3x3.3 mm; slice thickness, 3.0 mm; distance factor 0.17; flip angle 80°). In addition, a high-resolution T1-weighted magnetization-prepared rapid-acquisition gradient echo anatomical scan was obtained from each subject from a 3 Tesla MR scanner Magnetrom Trio Tim, Siemens Medical Systems, Erlangen, Germany) using a 32-channel head coil (192 sagittal slices; repetition time, 2.3s; echo time, 3.03ms; voxel size 1.0 x 1.0 x 1.0 mm; field of view 256 mm).

Preprocessing

fMRI data analysis was performed with SPM5 software (Statistical Parametric Mapping; Wellcome Trust Centre for Cognitive Neuroimaging, London, UK). The first 5 volumes of each participant's dataset were discarded to allow T1 equilibrium. First, realignment parameters were estimated for the images acquired at the first echotime and consequently applied to images resulting from the 3 other echoes. The echo-images were combined by applying a PAID-weight algorithm assessing the signal-to-noise ratio as described by Poser et al.(2006). Thirty volumes, acquired just after the main task (while the participant watched a black screen) were used as input for this algorithm. Thereafter the following preprocessing steps were applied: slice-time correction, co-registration and a segmentation procedure using the tissue probability maps provided by SPM5 for grey matter, white matter and CSF centered in MNI space to estimate normalization parameters based on the structural image. Structural as well as functional images were then normalized by applying these estimates. All normalized images were smoothed with an isotropic 8 mm full-width half-maximum Gaussian kernel(Worsley and Friston, 1995).

Single subject analysis

A random effects, event-related, statistical analysis was performed with SPM5. First, we specified a separate general linear model (GLM) for each participant. This GLM included 8 main regressors representing different factors of the MID task: Cue for reward and no-reward trials; Instrumental Target for reward and no-reward trials; Outcome for reward-hit, reward-miss, no-reward hit and no-reward miss trials. These regressors were modeled as delta functions at their onset and were convolved with a canonical hemodynamic response function. Realignment parameters (3 rigid-body translations and 3 rotations) were added to capture residual movement-related

artifacts. High-pass filtering (128s) was applied to remove low-frequency drifts. Parameter estimates for all regressors were obtained by maximum-likelihood estimation, modeling temporal autocorrelation (AR1). The parameter estimates, derived from this fit of the model to the data, reflect the strength of covariance between the fMRI data and the canonical response function for each of the regressors.

Group level analysis

The beta estimates for the Cue reward and no-reward trials were of primary interest. These were admitted to a 2x3 ANOVA (full factorial) at the group-level with Reward Expectation (reward/no-reward) as within-subject factor and Group (psychopathic criminals/noncriminal high-, noncriminal low- impulsive/antisocial) as between-subject factor. Restricted Maximum Likelihood estimates of variance components were used to allow for unequal variance between subjects and possible deviations from sphericity introduced by dependencies between levels in the repeated measures factor. For the factor Group between level independence was assumed. The main effects and interactions were then calculated.

We assessed whether increased signal during reward expectation relative to no-reward expectation was different between subjects high versus low on impulsive and antisocial traits as was expected based on Buckholz et al.(2010a). To this end, we used a planned contrast (within the full factorial, ANOVA model) to compare reward- (versus no-reward-)related signal change between the healthy controls with low impulsive/antisocial traits and the compound group with high impulsive/antisocial traits (i.e. healthy controls scoring high on these traits plus the psychopathic criminals who also scored high on these traits). Next we assessed whether reward- (versus no-reward-)related signal differed between the psychopathic criminals and the high impulsive/antisocial healthy controls (at the whole brain level as well as within our small volume of interest in the ventral striatum, family wise error corrected for multiple comparisons). Finally, we assessed group differences in reward-related signal by exploring the omnibus full factorial model. Supplementary, we repeated the full factorial ANOVA, but now with the healthy control group divided in two groups based on their Fearless/Dominance score (i.e. factor 1 of the PPI, see supplementary materials for results).

Functional connectivity analyses

Because we did not find differences between noncriminal controls with high impulsive/antisocial traits and psychopathic criminals on the contrasts described above, we anticipated that differences between criminal and noncriminal people with impulsive/antisocial traits might not lie in reward expectancy signals in the ventral striatum per se, but in how this region is connected to other brain regions. Therefore we assessed differences between these 2 groups in terms of functional connectivity (Friston et al., 1997) with the ventral striatal region, in which reward-related signal was increased in (criminals and noncriminal) people with high impulsive/antisocial traits versus people with low impulsive/antisocial traits. To conduct this analysis, we proceeded in several steps. First, for each individual in the psychopathic criminal and non-criminal impulsive/antisocial control group the (first principal component of the) BOLD time series was extracted from a 3 mm sphere surrounding the BOLD response peak revealed by contrast (1) (the seed, c.q. the right ventral striatum: Figure 6.3). The time series was then deconvolved based on the canonical hemodynamic response function (HRF) to construct a time series of neural responses following the procedures outlined by Gitelman et al.(2003).

Second, a GLM was estimated for every subject, which included the following 3 main regressors (as well as the 6 motion parameters): (1) The seed BOLD response time series; (2) a regressor representing the task-induced effect reflecting reward versus no-reward expectation; and (3) the psychophysiological interaction regressor, which is the cross product of the deconvoluted regressor (1) and regressor (2). The latter regressor represents the interaction between neural signal and the 2 task conditions. This regressor was then convolved with the HRF. Parameter estimates for the interaction regressor were estimated by maximum-likelihood estimation, modeling temporal autocorrelation at the subject-level. The parameter estimates, derived from this fit of the model to the data, reflect the strength of the task-induced change in connectivity with the seed region (the ventral striatum). These estimates were then used at the group level in an independent sample t-test with group (2 levels: psychopathic criminal/noncriminal impulsive/antisocial control) as between-subject factor to assess group differences.

Statistical thresholding of fMRI analysis

We report only those effects that survive family wise error correction for multiple comparisons at the whole brain ($p<.05$, voxel-level) and where

appropriate in the right ventral striatum as defined by the Harvard-Oxford Atlas (based on previous findings by Buckholtz et al.(2010a) and Bjork et al.(2012)).

6.3 RESULTS

6.3.1 *Behavioural results*

No behavioural differences between groups

During reward expectation subjects showed more accurate responses than during the neutral condition (main effect of Reward Expectation on hits: $F_{(1/31)}$ = 8.3, p = 0.007) leading to shorter target duration for the reward condition than the no-reward condition ($F_{(1/31)}$ = 11.0, p = 0.002). Critically, there were no main or interaction effects of Group in terms of target duration (F<1.5). Furthermore, as intended by our task design, behavioral performance on hits (i.e. accuracy) did not differ between the groups.

6.3.2 *Neuroimaging results*

First, analysis of data from all 34 subjects replicated prior studies using this task and revealed significantly greater BOLD signal in the ventral striatum during reward than no-reward cues (xyz = [12 14 -6], T=3.49, p =0.007; small volume correction for multiple comparisons within the anatomically defined right ventral striatum, (Knutson et al., 2003). Next, we established that, following Buckholtz et al.(Buckholtz et al., 2010a), reward-related BOLD signal in the ventral striatum was higher in the 2 groups with high impulsive/antisocial traits than in the group with low impulsive/antisocial traits (Figure 6.3, whole-brain corrected for multiple comparisons: T=5.31, p=0.049, small volume correction for ventral striatum: T=3.30, p=0.011). There were no other Group x Reward effects (established by an omnibus ANOVA with a 3-level group factor, and/or planned contrasts between pairs of groups). Critically, there were also no differences between the 2 (criminal versus noncriminal) groups with high impulsive/antisocial traits. Thus reward-related BOLD signal in the ventral striatum was enhanced in people with impulsive/antisocial traits, but did not differentiate criminal psychopathic individuals from noncriminal individuals with impulsive/antisocial traits. Furthermore, there were no correlations between task performance (reaction times/target duration on reward versus no-reward trials) and BOLD signal change (reward – no reward) in the ventral striatum.

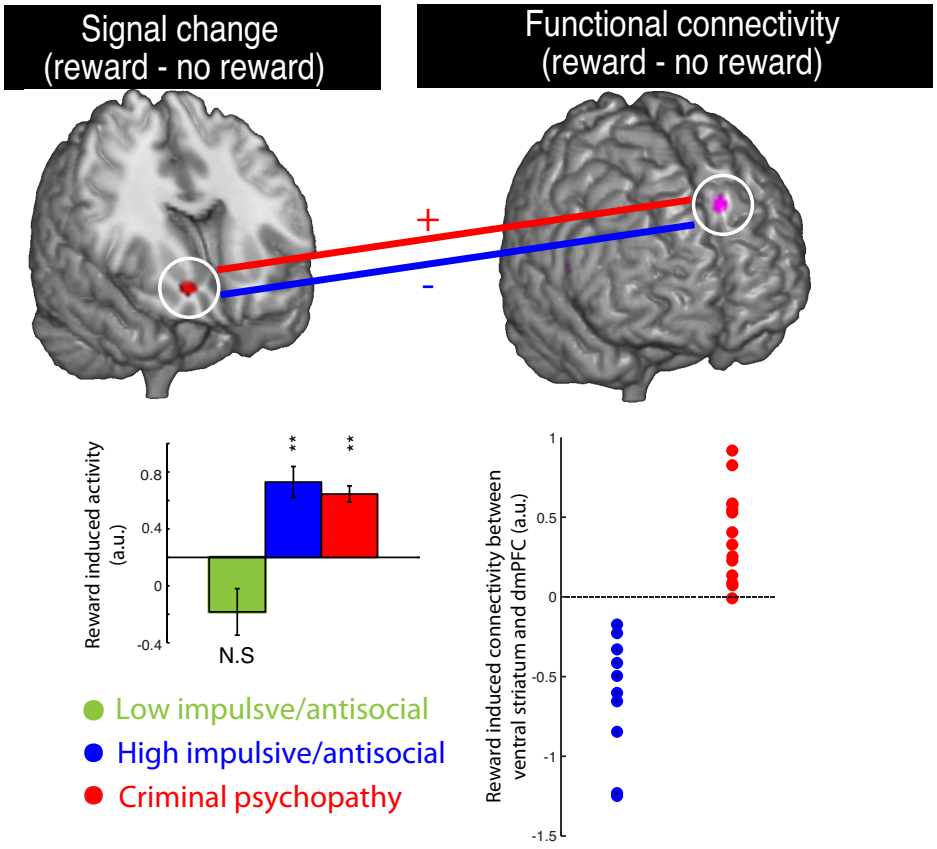


Figure 6.3 *– Enhanced reward-related BOLD signal in the ventral striatum of psychopathic criminals and healthy noncriminal controls with high antisocial/impulsive traits compared with healthy controls with low antisocial/impulsive traits (peak: MNI XYZ [18 22 -8]). Average signal change (reward – no-reward) extracted from the peak cluster is shown for illustrative purposes. (b) BOLD signal in the ventral striatum (3 mm sphere around MNI XYZ [18 22 -8]) contributes differentially to the dorsomedial prefrontal cortex (dmPFC) during reward versus no reward expectancy in psychopathic criminals compared with healthy controls with high antisocial/ impulsive traits. The scatter plot depicts individual parameter estimates of functional connectivity differences between reward and no reward expectation, extracted from the peak cluster (peak: MNI XYZ [-14 34 44)]). Images are displayed at a threshold of p<0.001 uncorrected for illustration purposes.*

Next we tested the hypothesis that the difference between the noncriminal impulsive/antisocial group and psychopathic criminal group does not lie in ventral striatal signaling per se, but rather in the degree to which this reward-related neural signal interacts with neural systems that control behavior, such as the prefrontal cortex. This task-dependent, functional connectivity analysis revealed a group effect in the dorsomedial prefrontal cortex. No other regions were revealed by this analysis. Thus, reward-related connectivity between the ventral striatum and dorsomedial prefrontal cortex was different between the psychopathic criminals and the noncriminal impulsive/antisocial group (Figure 6.3, 2 sample t-test: T=7.44, p=0.018, result was corrected for multiple comparisons at the whole brain level). This difference was remarkable in terms of consistency (Figure 6.3): There was no overlap between the groups, with reward-related connectivity between the ventral striatum and dorsomedial prefrontal cortex being below zero in all healthy high impulsive/antisocial individuals, but (around or) above zero in all psychopathic criminals. There were no significant correlations between connectivity differences (reward versus no-reward) within the psychopathy group and the factors of the two or four factor model of the PCL-R (within the psychopathy group, all Spearman's rho<.580, p>.007, p-threshold corrected for multiple comparisons). Moreover, there were no correlations between task performance (reaction times/target durations on reward versus no-reward trials) and connectivity (reward versus no-reward)..

6.4 DISCUSSION

The present data suggest that not reward expectation per se, but the way in which reward expectations are communicated to frontal areas might be key to understanding the overt criminality in impulsive/antisocial people. This suggests that criminality in impulsive/antisocial individuals is accompanied by abnormal contribution of reward signaling to regions regulating the cognitive control of behavior (Ridderinkhof, 2004).

We go beyond earlier studies that assessed reward expectation in relation to impulsive/antisocial traits (Buckholtz et al., 2010a; Bjork et al., 2012) by assessing high impulsive/antisocial criminals (compared with healthy control samples) and by assessing task-related connectivity. Our results show differential task-dependent coupling between the anterior ventral striatum and dorsomedial prefrontal cortex (c.q. superior frontal gyrus) in the high impulsive/antisocial healthy control group compared with the high impulsive/antisocial criminals. Noteworthy is the difference between these groups, which was striking in terms of its nature and robustness: First, there

was no overlap between the groups when assessing individual connectivity patterns. Second, all impulsive/antisocial healthy controls showed a clear negative coupling, whereas all the criminal individuals showed (near zero or) positive coupling.

These results advance findings from earlier studies showing abnormal reward processing in healthy, non-criminal volunteers with impulsive/antisocial traits (Buckholtz et al., 2010a; Bjork et al., 2012) to a criminal sample. For example, Buckholtz et al. (2010a) have reported a positive association between impulsive/antisocial trait scores and neural signal in the right ventral striatum during reward expectation. The ventral striatum well connected with (para)limbic, cortical and ventral tegmental areas as well as motor effector sides, which enables this region to functions as an interface between cognition, emotion and action(Cardinal et al., 2002; Floresco, 2015). It has an established function in reward-related processes such as expectation of reward(Knutson et al., 2001) and ventral striatal deficits may be involved in impulsivity(Basar et al., 2010), sensation seeking and heightened reward sensitivity all associated with antisocial behaviour (Glenn and Yang, 2012). Recently, greater reactivity in the ventral striatum has been directly linked with increased retaliatory aggression (Chester and DeWall, 2015). Moreover, there is one case description of deep brain-stimulation in bilateral ventral striatum resolving pathological (self-directed) aggression(Harat et al., 2015). Furthermore, the finding that enhanced reward signaling in the ventral striatum is not specific to impulsive/antisocial criminals, but extends to noncriminal, but impulsive/antisocial individuals is not surprising given previous studies showing enhanced reward signaling in the ventral striatum of healthy individuals with high impulsive(Buckholtz et al., 2010b) or impulsive/antisocial traits (Buckholtz et al., 2010a).

Our findings concur with a growing body of research that suggests that subjects comparable in terms of overt criminality and PCL-score to our sample (i.e. psychopathic criminals) are characterized by aberrant connectivity within networks that underpin the interaction between affective and cognitive processes (Yang et al., 2012; Contreras-Rodríguez et al., 2014; Motzkin et al., n.d.). In fact psychopathic criminals have been shown to exhibit abnormal functional and structural connectivity patterns in particular with the dorsomedial prefrontal cortex (Yang et al., 2012; Contreras-Rodríguez et al., 2014). This part of the prefrontal cortex has long been implicated in the cognitive control of behaviour, especially in signalling the need for performance adjustment(e.g. Ridderinkhof, 2004), self-inhibition of movements (Brass and Haggard, 2007) and impulse control(Cho et al., 2013). Interestingly, non-invasive stimulation of the dorsomedial prefrontal cor-

tex via repetitive transcranial magnetic stimulation has been shown to enhance inhibitory control over prepotent responses(Obeso et al., 2013), to improve subjective choice for delayed rewards, and to interfere with striatal dopamine(Cho et al., 2015). These observations raise the hypothesis that psychopathic criminals might exhibit a failure to adjust performance due to aberrant impact of reward expectation. This hypothesis should be tested in future studies with behavioural tasks that are optimized for detecting aberrant behaviour in psychopathic criminals. We hypothesize that tasks involving both reward anticipation and adjustment of behaviour based on punishment and/or of negative (facial) emotional cues (cf. Blair, 2008) might be particularly sensitive.

The present results might be relevant in the context of theorizing about differences between "successful" from "unsuccessful" psychopathic individuals, with the former referring to individuals with psychopathic personality traits who do not have any criminal convictions(Gao and Raine, 2010). Unlike previous studies that focused on reward expectation in healthy people with psychopathic traits(Buckholtz et al., 2010a; Bjork et al., 2012), the present study included also a sample of criminal, non-successful psychopathic individuals rather than a community sample of people with high PPI scores. As such our data concur with previous suggestions that there are substantial differences between non-successful and successful psychopathic individuals at neural, physiological, cognitive and behavioral levels(Gao and Raine, 2010). Specifically, non-successful psychopathic individuals have been argued to exhibit greater frontal impairments and greater high-level cognitive deficits than do successful psychopathic individuals(Gao and Raine, 2010). Our results provide the first direct evidence for this hypothesis. Note that the comparison between successful and non-successful impulsive/antisocial individuals is necessarily confounded by overt criminal history. As such, we cannot and do not claim specificity of our findings to psychopathic criminals compared with non-psychopathic criminals.

Previous studies have shown enhanced reward expectancy signaling in the ventral striatum of healthy individuals with high impulsive(Plichta and Scheres, 2014) or impulsive/antisocial traits(Buckholtz et al., 2010a). Nevertheless, it is far from trivial that psychopathic criminals show similar hypersensitivity, because several other patient groups characterized by impulsivity, such as ADHD, show reduced ventral striatal neural signals during reward expectancy(Plichta and Scheres, 2014). The present results contribute to our understanding of why some people act according to their impulsive/antisocial personality while others are able to behave adaptively despite reward-related urges. The enhanced reward expectation processing

in psychopathic criminals concurs very well with the clinical observation that psychopathic criminals exhibit ruthless reward-driven behaviors and that they, unlike healthy people, are not inhibited by immoral or otherwise aversive signals. Moreover, this account would accord with the finding that psychopathic criminals fail to adapt behavior based on aversive information when reward is at stake(Newman et al., 1990).

It is important to emphasize that our findings cannot be interpreted as being specific to either criminality, psychopathy or their interaction (i.e. psychopathic criminality): a non-criminal control group with high psychopathy severity (in terms of PCL-R scores) nor a criminal control group with low psychopathy severity was included. Crucially, our findings enhance the understanding of the neural mechanism underlying overt criminality within impulsive/antisocial populations, given the comparison of two (one criminal, one non-criminal) equally impulsive/antisocial groups.

Furthermore, we highlight that the sample size in this study was relatively modest, which is a necessary consequence of the limited availability of clean, drug- and tattoo-free psychopathic criminals. Indeed our sample size is comparable with that in other fMRI studies with psychopathic criminals. Nevertheless, we argue that our results are reliable, given that it includes a general replication of prior findings(Buckholtz et al., 2010a), and given that our novel result on connectivity reaches statistical significance even after correction for multiple comparisons at the whole-brain level. As such, the present mechanistic study raises a promising target for future clinical work, which could advance our findings to a diagnostic level, by replicating in a larger population our finding that reward expectancy-based connectivity fully dissociated criminals from high impulsive/antisocial non-criminals.

6.5 REFERENCES

BASAR K, Sesia T, Groenewegen H, Steinbusch HWM, Visser-Vandewalle V, Temel Y (2010) Progress in Neurobiology. Progress in Neurobiology 92:533–557.

BJORK JM, Chen G, Hommer DW (2012) Psychopathic tendencies and mesolimbic recruitment by cues for instrumental and passively obtained rewards. Biological Psychology 89:408–415.

BLAIR RJR (2008) The amygdala and ventromedial prefrontal cortex: functional contributions and dysfunction in psychopathy. Philosophical Transactions of the Royal Society B: Biological Sciences 363:2557–2565.

BLAIS J, Solodukhin E, Forth AE (2014) A meta-analysis exploring the relationship between psychopathy and instrumental versus reactive violence. Criminal Justice and behavior, 41, 707-821.

BRASS M, Haggard P (2007) To do or not to do: the neural signature of self-control. Journal of Neuroscience 27:9141–9145.

BUCKHOLTZ JW, Treadway MT, Cowan RL, Woodward ND, Benning SD, Li R, Ansari MS, Baldwin RM, Schwartzman AN, Shelby ES, Smith CE, Cole D, Kessler RM, Zald DH (2010a) Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. Nature Neuroscience 13:419–421.

BUCKHOLTZ JW, Treadway MT, Cowan RL, Woodward ND, Li R, Ansari MS, Baldwin RM, Schwartzman AN, Shelby ES, Smith CE, Kessler RM, Zald DH (2010b) Dopaminergic network differences in human impulsivity. Science 329:532.

CAMP JP, Skeem JL, Barchard K, Lilienfeld SO, Poythress NG (2013) Psychopathic predators? Getting specific about the relation between psychopathy and violence. Journal of Consulting and Clinical Psychology 81:467–480.

CARDINAL RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience and Biobehavioral Reviews 26:321–352.

CHESTER DS, DeWall CN (2015) The pleasure of revenge: retaliatory aggression arises from a neural imbalance toward reward. Social Cognitive and Affective Neuroscience, nsv082, Doi: 10.1093/scan/nsv082.

CHO SS, Koshimori Y, Aminian K, Obeso I, Rusjan P, Lang AE, Daskalakis ZJ, Houle S, Strafella AP (2015) Investing in the future: stimulation of the medial prefrontal cortex reduces discounting of delayed rewards. Neuropsychopharmacology 40:546–553.

CHO SS, Pellecchia G, Aminian K, Ray N, Segura B, Obeso I, Strafella AP (2013) Morphometric correlation of impulsivity in medial prefrontal cortex. Brain Topogr 26:479–487.

CONTRERAS-RODRÍGUEZ O, Pujol J, Batalla I, Harrison BJ, Soriano-Mas C, Deus J, López-Solà M, Macià D, Pera V, Hernández-Ribas R, Pifarré J, Menchón JM, Cardoner N (2014) Functional Connectivity Bias in the Prefrontal Cortex of Psychopaths. Biological Psychiatry, 78(9), 647–55.

EDENS JF, Poythress NG, Lilienfeld SO, Patrick CJ, Test A (2008) Further evidence of the divergent correlates of the Psychopathic Personality Inventory factors: Prediction of institutional misconduct among male prisoners. Psychological Assessment 20:86–91.

FLORESCO SB (2015) The Nucleus Accumbens: An Interface Between Cognition, Emotion, and Action. Annu Rev Psychol 66:25–52.

FRISTON KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6:218–229.

GAO Y, Raine A (2010) Successful and unsuccessful psychopaths: a neurobiological model. Behav Sci Law 28:194–210.

GITELMAN DR, Penny WD, Ashburner J, Friston KJ (2003) Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. NeuroImage 19:200–207.

GLENN AL, Yang Y (2012) The potential role of the striatum in antisocial behavior and psychopathy. Biological Psychiatry 72:817–822.

HARAT M, Rudas M, Zielinski P, Birska J, Sokal P (2015) Deep Brain Stimulation in Pathological Aggression. Stereotact Funct Neurosurg 93:310–315.

HARE RD (2003) Manual for the Hare Psychopathy Checklist-Revised, 2nd ed. Toronto, Ontario, Canada: Multi Health Systems.

HARE RD, Neumann CS (2008) Psychopathy as a Clinical and Empirical Construct. Annu Rev Clin Psychol 4:217–246.

HOOGMAN M, Onnink M, Cools R, Aarts E, Kan C, Arias Vasquez A, Buitelaar J, Franke B (2013) The dopamine transporter haplotype and reward-related striatal responses in adult ADHD. Eur Neuropsychopharmacol 23:469–478.

JELICIC M, Merckelbach H, Timmermans M, Candel I (2004) De Nederlandstalige versie van de Psychopathic Personality Inventory. Psychodiagnostisch gereedschap. De Psycholoog:604–608.

KENNEALY PJ, Skeem JL, Walters GD, Camp J (2010) Do core interpersonal and affective traits of PCL-R psychopathy interact with antisocial behavior and disinhibition to predict violence? Psychological Assessment 22:569–580.

KNUTSON B, Adams CM, Fong GW, Hommer D (2001) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. Journal of Neuroscience 21:RC159.

KNUTSON B, Fong GW, Bennett SM, Adams CM, Hommer D (2003) A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. NeuroImage 18:263–272.

LILIENFELD SO, Andrews BP (1996) Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. J Pers Assess 66:488–524.

MOTZKIN JC, Newman JP, Kiehl KA, Koenigs M (n.d.) Reduced Prefrontal Connectivity in Psychopathy, Journal of Neuroscience, 31(48), 17348–57.

NEUMANN CS, Vitacco MJ, Hare RD, Wupperman P (2005) Reconstruing the "reconstruction" of psychopathy: a comment on Cooke, Michie, Hart, and Clark. J Pers Disord 19:624–640.

NEWMAN JP, Patterson CM, Howland EW, Nichols SL (1990) Passive avoidance in psychopaths: The effects of reward. Personality and Individual Differences 11:1101–1114.

OBESO I, Cho SS, Antonelli F, Houle S, Jahanshahi M, Ko JH, Strafella AP (2013) Stimulation of the pre-SMA influences cerebral blood flow in frontal areas involved with inhibitory control of action. Brain Stimul 6:769–776.

PATRICK CJ, Fowles DC, Krueger RF (2009) Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. Develop Psychopathol 21:913.

PLICHTA MM, Scheres A (2014) Neuroscience and Biobehavioral Reviews. Neuroscience and Biobehavioral Reviews 38:125–134.

PORTER S, Woodworth M (2006) Psychopathy and Aggression. In: Handbook of psychopathy, 1st ed. (Patrick CJ, ed), pp 481–494. New York: The Guilford Press.

POSER BA, Versluis MJ, Hoogduin JM, Norris DG (2006) BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. Magn Reson Med 55:1227–1235.

RIDDERINKHOF KR (2004) The Role of the Medial Frontal Cortex in Cognitive Control. Science 306:443–447.

SCHMAND B, Bakker D, Saan R, Louman J (1991) [The Dutch Reading Test for Adults: a measure of premorbid intelligence level]. Tijdschrift voor Gerontologie en Geriatrie 22:15–19.

VAN der Gronde T, Kempes M, van El C, Rinne T (2014) Neurobiological Correlates in Forensic Assessment: A Systematic Review. PLoS ONE, 9(10).

VAN Vliet IM, Leroy H, van Megen HJGM (n.d.) M.I.N.I Internationaal Neuropsychiatrisch Interview, 5 ed.

WEERTMAN A, Arntz A, Kerkhofs M (2000) Handleiding gestructureerd klinisch interview voor DSM-IV AS-II persoonlijkheidsstoornissen. Lisse, The Netherlands: Swets test publishers.

WHITTINGTON R, Hockenhull JC, McGuire J, Leitner M, Barr W, Cherry MG, Flentje R, Quinn B, Dundar Y, Dickson R (2013) A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour: update 2002–8. Health Technology Assessment 17.

WICKRAMASEKERA N, Wright J, Elsey H, Murray J, Tubeuf S (2015) Cost of crime: A systematic review. Journal of Criminal Justice 43:218–228.

WORSLEY KJ, Friston KJ (1995) Analysis of fMRI time-series revisited--again. NeuroImage 2:173–181.

YANG Y, Raine A, Joshi AA, Joshi S, Chang YT, Schug RA, Wheland D, Leahy R, Narr KL (2012) Frontal information flow and connectivity in psychopathy. The British Journal of Psychiatry 201:408–409.

6.6 SUPPLEMENT

In response to helpful suggestions we got through the peer-review process, we post hoc conducted the same analysis as we presented in the paper, but now focussed on PPI factor 1 (fearless/dominance) scores instead of PPI factor 2 (impulsive/antisociality) scores. Thus, we formed three groups: (1) healthy controls scoring low and (2) high on fearless/dominance and (3) psychopathic criminals (scoring equally high on fearless/dominance as the high scoring healthy controls: mean (SD): high scoring controls: 153.3(14.6), psychopathic criminals: 142.6(19.9), 2 sample t-test: $t_{(22)}$=1.44, p=.17). Again, the beta estimates for the Cue reward and no-reward trials were admitted

to a 2x3 ANOVA (full factorial) at the group-level with Reward Expectation (reward/no-reward) as within-subject factor and Group (psychopathic criminals/noncriminal high-, noncriminal low-fearless/dominance) as between-subject factor. We again assessed whether there was an interaction between Group and Reward Expectancy.

The rmANOVA revealed an interaction of Group (3 levels) and Reward Expectancy (2 levels) in the peri-aqueductal grey (PAG) (F-test, whole-brain corrected for multiple comparisons: xyz = [2 -32 -6], Z=5.09, p =0.017, see Supplementary figure 6.4). This effect was driven by differential BOLD response to the reward cue (compared to the neutral cue) in high fearless/-dominant controls compared to the psychopathic criminal group (T-test, whole-brain corrected for multiple comparisons: xyz = [2 -32 -6], Z=5.56, p=.001).

Reactivity of the PAG in humans has been related to proximity of threat (Mobbs, 2007) and reactive aggression (Blair, 2015). Of specific interest in this perspective is a recent experiment by Yu et al. (2014), that assessed frustration by blocking expected reward outcomes. They showed that blocking an expected reward led to significant brain responses in the PAG at the time of blocking. Moreover, the response of the PAG was also modulated by the proximity of the blocked reward and by the expended effort till the point of blocking. Here we show that overt (psychopathic) criminals differ from non-criminals who also score high on fearless/dominance traits in PAG reactivity when a cue is signaling whether they can or cannot (cf. blocking) obtain a reward. On a speculative account these findings suggest that overt criminality in fearless/dominant subjects might be accompanied by a differential PAG response to frustration in the context of reward expectation (i.e. the blocking an expected reward).

6.6.1 *References*

BLAIR, R. J. R. (2015). The Neurobiology of Impulsive Aggression. Journal of Child and Adolescent Psychopharmacology, cap.2015.0088. http://doi.org/10.1089/cap.2015.0088

MOBBS, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., et al. (2007). When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. Science, 317(5841), 1079–1083. http://doi.org/10.1126/science.1144298

YU, R., Mobbs, D., Ben Seymour, Rowe, J. B., & Calder, A. J. (2014). ScienceDirect. Cortex, 54(C), 165–178. http://doi.org/10.1016/j.cortex.2014.02.013

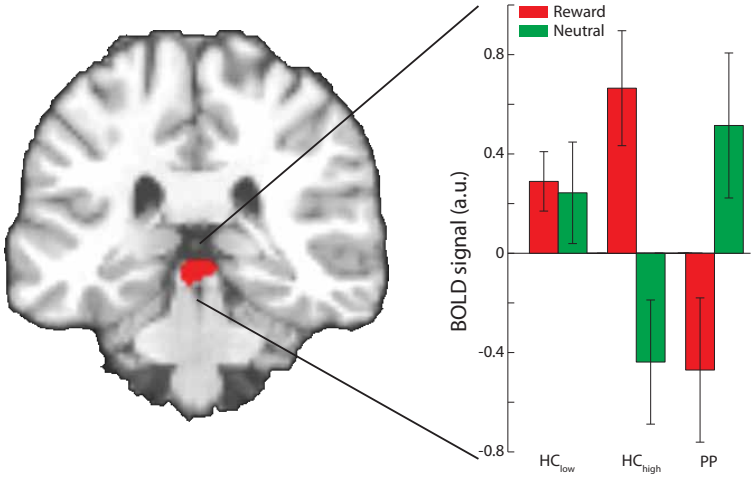


Figure 6.4 *– Interaction between Group (3 levels: non criminals scoring (i) low (HC_{low}) and (ii) high (HC_{high}) on the PPI Fearless/Dominance factor and (iii) psychopathic criminals) and Reward Expectancy (2 levels: (reward/neutral) in the peri-aquaductal grey (whole-brain corrected for multiple comparisons: xyz = [2 -32 -6], Z=5.09, p =0.017). Brain image is displayed at a threshold of p<0.001 uncorrected for illustration purposes. Bar graph displays the average beta estimates extracted from the PAG cluster for illustrative purposes.*

Part IV

CLOSING

7

SUMMARY AND DISCUSSION

This thesis described a collection of studies assessing the motivational control of instrumental behaviour by affective cues. In this general discussion I will first give a brief summary of the results of the different studies. I will then synthesize what insights may be gained from these studies when considered together. I will describe these insights respectively related to the motivational influence of affect on behaviour, the neurocognitive underpinnings of this motivational influence, the role of serotonin and how these go awry in two psychiatric disorders. Next, limitations of the findings in this thesis will be discussed. Finally (and throughout this discussion), I will discuss perspectives on future studies.

7.1 SUMMARY

In chapter 1 I described the aim of this thesis and its background. The aim of the studies presented in this thesis was (i) to increase our understanding of the neural and neurochemical mechanisms that allow affective states to alter instrumental actions in healthy humans; and (ii) to use this knowledge to leverage our understanding of the cognitive and neural mechanisms underlying affective dysregulation of behaviour in patients with affective, impulsive and aggressive symptoms. In chapter 1 I argued that the general Pavlovian to instrumental transfer (PIT) paradigm provides a promising operationalization of the motivational control of instrumental behaviour by affective cues.

In chapter 2 I investigated the role of serotonergic transmission in PIT. Serotonin had been implicated in a wide range of healthy functions and disordered behaviour, including psychiatric symptoms. This neurotransmitter had been hypothesized to couple punishment to behavioural inhibition (Dayan and Huys, 2008; 2009; Boureau and Dayan, 2011; Cools et al., 2011). The work presented in chapter 2 was the first to assess this suggested effect of punishment on action via serotonergic transmission in a paradigm that directly assessed the influence of affective stimuli on independently acquired instrumental behaviour. Specifically, I investigated the role of serotonin in how affective states (de)motivate instrumental actions across aversive and appetitive domains. Attenuation of serotoninergic transmission was indeed accompanied by attenuation of the inhibiting effect of specifically aversive Pavlovian cues on instrumental behaviour. This finding supported theories that hold that intact serotonergic transmission is necessary for coupling aversive expectations to behavioural inhibition (Dayan and Huys, 2008; 2009; Boureau and Dayan, 2011; Cools et al., 2011).

Next, in chapter 3 I assessed the neural underpinnings of this interaction with a PIT task specifically adjusted for fMRI analyses. I showed that BOLD responses in the amygdala and the ventral striatum were associated with behavioural inhibition by aversive Pavlovian cues across approach and withdrawal actions. Furthermore, BOLD responses in the ventromedial prefrontal cortex (vmPFC) differed between approach and withdrawal actions. Aversive affective stimuli modulated connectivity between the vmPFC and the ventral and ventromedial striatum. These results show that action-specific aversive control of instrumental behaviour involves the modulation of fronto-striatal interactions by affective stimuli.

In the second part of this thesis I investigated the translational potential of the PIT paradigm for psychiatry, focusing on two patient groups: Patients

with borderline personality disorder (chapter 4) and patients with psychopathy (chapter 5 and 6). Aberrant affective processing and behavioural disturbances are thought to be key for both disorders (Linehan, 1993; Rosenthal et al., 2008; R Blair, 2013). Behaviour of psychopathic criminals is characterized by cold, instrumental behaviour especially targeted at their own comfort and hardly disturbed by discomfort of others, which is usually aversive to non-psychopathic people. As such, we hypothesized that their behaviour might be characterized by too little behavioural inhibition in the face of aversive information, as well as by behaviour that is overly targeted at personal gains. Patients with borderline personality disorder suffer from emotional instability and their behaviour often seems overly reactive to affective influences. In both disorders I assessed whether there was an association between the motivational influence of affect on instrumental behavior and associated neural responses on the one hand and specific clinical measures on the other hand. The clinical measure of primary interest in the psychopathy study was psychopathy severity (measured with the psychopathy checklist revised (PCL-R)). In the study with borderline personality disorder I focused on borderline personality symptom severity (measured with the borderline personality disorder checklist (BPD47)) and changes in this severity 1 year after the start of treatment.

In chapter 4 I found that BOLD signal in the bilateral amygdala during aversive PIT predicted symptom reduction at one year follow-up: Increased aversive PIT in amygdala signaling (i.e. CS-depended coupling between BOLD-signal and instrumental behaviour) before treatment was associated with less clinical improvement after treatment. Thus, long-term clinical symptom reduction in BPD patients could be predicted from BOLD signal in the amygdala related to the motivational interaction between affect and instrumental behaviour. This finding demonstrates a key role for abnormal aversive processing in the amygdala in borderline personality disorder and begins to elucidate the factors that predict recovery from this disorder.

In the psychopathy PIT study (chapter 5) I found that psychopathy severity postively correlated with an attenuation of inhibition of instrumental action by aversive affective stimuli. Moreover, there was an anomalous positive association between aversive inhibition of action and aversive inhibition of BOLD signal in the caudate nucleus in psychopathic criminals. These results show that psychopathy severity is associated with reduced transfer of aversive Pavlovian motivation to instrumental action inhibition. In addition our results suggest that this aversive Pavlovian inhibition might be due to inappropriate transfer of aversive Pavlovian values to neural systems involving the caudate nucleus.

To additionally address the neural effects of reward in the sample of psychopathic criminals I assessed the neural correlates of reward expectation in chapter 6. This research was pertinent because recent work (Buckholtz et al., 2010) had suggested a central role for reward expectation processing in impulsive/antisocial traits, which are central to the construct of psychopathy and relevant in predicting criminal behaviour. I thus compared the group of psychopathic criminals performing a reward anticipation task in the MRI scanner, to two groups of non-criminal individuals: One with high levels and another with low levels of impulsive/antisocial traits. Psychopathic criminals showed heightened neural reactivity in the ventral striatum in response to reward expectation. However, comparable, heightened reactivity was also found in non-criminal individuals with comparable (high) levels of impulsive/antisocial traits. Critically, psychopathic criminals differed from these non-criminal individuals in the communication between this hyperresponsive reward-related ventral striatal region and the dorsomedial prefrontal cortex: Psychopathic criminals showed positive reward-related connectivity between these regions in contrast to the non-criminal controls with high levels of impulsive/antisocial traits. Thus, incarcerated psychopathic criminals could be dissociated from non-criminal individuals with comparable impulsive/antisocial personality tendencies based on the degree to which reward-related brain regions interact with brain regions that control behaviour, such as the dorsomedial prefrontal cortex. These results may help us understand why some people act according to their impulsive/antisocial personality while others are able to behave adaptively despite reward-related urges.

Taken together, we showed that:

- Serotonin is involved in the motivational influence of aversive affective stimuli on instrumental behaviour.
- The amygdala and fronto-striatal circuitry are involved in the motivational control of instrumental behavior by aversive cues.
- Reduced aversive PIT BOLD signal in the amygdala predicts increased clinical symptom reduction one year after the start of treatment in patients with borderline personality disorder.
- Psychopathy severity within a group of psychopathic criminals is related to the level of behavioural inhibition by aversive stimuli and this might reflect aberrant recruitment of the caudate nucleus.

- Neural processing of reward expectation differentiated psychopathic criminals from non-criminals with similar impulsive/aggressive personality traits.

7.2 AFFECTIVE MOTIVATION OF BEHAVIOUR

As described throughout this thesis and detailed in the introduction, we can discern two major controllers of human behaviour: First, Pavlovian control arising from stimulus - outcome associations. Second, instrumental control of behaviour that arises from stimulus-action-outcome contingencies. Through associations formed during Pavlovian learning, a previously insignificant stimulus is laden with the value associated with the predicted outcome. The crucial finding on which this thesis built is that such a Pavlovian stimulus (laden with incentive value) can motivate or demotivate instrumental actions (e.g. Huys et al., 2011). This observation is replicated here in several independent studies in humans. Substantiating this observation in humans was pertinent: This phenomenon has been suggested to be at the core of irrational behaviours that constitute general biases in human behaviour (see introduction). Moreover, it has been associated with psychiatric pathologies and might be pivotal for a comprehensive understanding of these pathologies (Dayan and Huys, 2008; 2009; Huys et al., 2012; Dayan and Seymour, 2013; Dolan and Dayan, 2013; e.g. Heinz et al., 2016).

7.2.1 *Action specificity*

The main behavioural result replicated throughout the present studies is that aversive Pavlovian CSs inhibit instrumental approach behaviour (chapter 2-5), but promote instrumental withdrawal behaviour (chapter 3-5). Thus, the observed Pavlovian influences were action specific: The same CS had opposite effects on instrumentally acquired approach and withdrawal actions. This action specificity is not readily explained from existing theoretical models of PIT. How can the same Pavlovian CS exert opposite motivational influence (inhibition vs. activation) on different forms of instrumental behaviour?

This instantiation of action-specificity of the PIT effect on the motivational level requires that at some point the motivational value of the CS has to be evaluated against the nature of the instrumental action. This evaluation then has to result in directing the motivational value to either invigorate or inhibit the instrumental action. Such an evaluation has not been implemented in any of the cognitive explanatory models of PIT (Rescorla and

Solomon, 1967; Dickinson and Balleine, 2002; Balleine and Ostlund, 2007; Cartoni et al., 2013; Cohen-Hatton et al., 2013). This incorporation would be interesting in its own account, but is particularly pressing because action-specificity of PIT has recently been shown to be predictive of recovery from depression (Huys et al., 2016).

7.2.2 *Affective dysregulation of choice and vigour: a simple meta-analysis*

I was unable to show significant (action-specific) PIT effects in terms of choice or vigour in some of the presented studies. This might reflect a lack of power due to the relatively small sample sizes. Here I address this issue by combining the studies that employed the same PIT-paradigm (chapter 3 (n=33), 4 (n=33) and 5 n=(31)) in one meta-analysis with considerable sample size (n=97).

In terms of choice (go/nogo), the overall rmANOVA (Study Group (6 levels) x Action Context (approach/withdrawal) x CS Valence (appetitive/neutral/aversive) reveals a strong overall Action Context x CS Valence interaction ($F(1.485, 135.1)=11.5$, $p<.001$, GC corrected, $\eta^2_p=.112$), with strong evidence for Action Context specific aversive PIT ($F(1,91)=17.0$, $p<.001$, $\eta^2_p=.158$) and no evidence for Action Context specific appetitive PIT ($F(1,91)=2.0$, $p=.165$, $\eta^2_p= .021$). This interaction was due to aversive suppression of approach responses as well as aversive enhancement of withdrawal responses. Simple effects analyses revealed significant aversive PIT effects for approach ($F(1,91)=9.1$, $p=.003$, $\eta^2_p=.091$) and withdrawal ($F(1,91)=12.8$, $p=.001$, $\eta^2_p=.123$). Thus I find strong evidence for action specific aversive PIT in terms of go/nogo choice.

Similar and additional effects were observed for vigour (button presses) (main effect of CS Valence: $F(1.8,165.0)=4.2$, $p=.020$, $\eta^2_p=.044$; Action Context x CS Valence: $F(1.7,157.3)=3.6$, $p=.037$, $\eta^2_p=.038$, GG corrected). Assessment of each Action Context separately revealed that there was a main effect of CS Valence in approach ($F(1.67,152.0)=7.4$, $p=.002$, $\eta^2_p=.075$, GG-corrected), but not in withdrawal ($F(1.84,167.6)=0.8$, $p= .456$, $\eta^2_p=.009$). The effect of CS Valence in approach was driven by aversive PIT (neutral versus aversive: $F(1,91)=5.1$, $p=.011$, $\eta^2_p=.069$), but not by appetitive PIT (neutral versus appetitive: $F(1,91)=1.7$, $p=.194$, $\eta^2_p=.018$).

A sample size calculation for a repeated measures design (http://glimmpse.samplesizeshop.org/) based on this meta-analysis suggests that a sample of 35 participants would be needed for detecting action specific aversive PIT in terms of choice (for a power of 0.9, alpha level 0.05 and based on the cell-means, standard deviations and correlations between measure-

ments derived from the above meta-analysis). The study groups presented in this thesis consisted of considerable smaller samples. The small samples size might then also explain why I was not able to find any behavioural differences between groups.

7.3 NEURAL CIRCUITRY OF MOTIVATION: THE CASE OF AVERSIVE PIT

In contrast to appetitive PIT (Balleine and Doherty, 2009) there is no comparable neurocognitive working model of aversive PIT (but see McDannald, 2009). I propose a putative neural model for action-specific aversive PIT in humans. This model is schematically summarized in Figure 7.1. In this model I hypothesize that the amygdala is primarily involved in valuating the presented CS and feed this information forward to the (ventral) striatum, where instrumental and affective information are integrated. In addition, I propose that action-specificity of the PIT effect is implemented by the vmPFC through amygdala-vmPFC and vmPFC-striatal connectivity.

Greater aversive behavioral PIT effects were related to enhanced activation of the amygdala (chapter 3), ventral striatum (chapter 3) and putamen (chapter 5) during aversive CS presentation. These findings generally concur with studies on appetitive PIT in humans, showing that greater appetitive PIT effects were related to enhanced activation of the amygdala (Talmi et al., 2008; Prevost et al., 2012), ventral striatum (Talmi et al., 2008), and putamen (Mendelsohn et al., 2014) during CS presentation. These converging findings show that the more BOLD response is evoked by a CS+ across trials in the amygdala and (ventral) striatum, the stronger the average PIT effect. These findings in humans together with the findings in animals (see Introduction) suggest that the amygdala, ventral striatum and putamen are necessary for generating PIT.

Note that these previous findings are based on between-subject analysis. In addition, two human studies found that ventral striatum was also involved in the integration of instrumental action and CS valence on a trial-by-trial basis: The study by Talmi and colleagues (2008) showed this for appetitive PIT and that by Garbusow (2016) and colleagues showed this across appetitive and aversive PIT trials. Thus, the ventral striatum might serve in PIT as the limbic-motor interface (Mogenson et al., 1980), integrating Pavlovian motivation with instrumental action, putatively through input from the amygdala.

At this moment only indirect evidence exists to hypothesize a connection between the amygdala and ventral striatum in PIT either directly via the basolateral amygdala or indirectly via the ventral tegmental area and

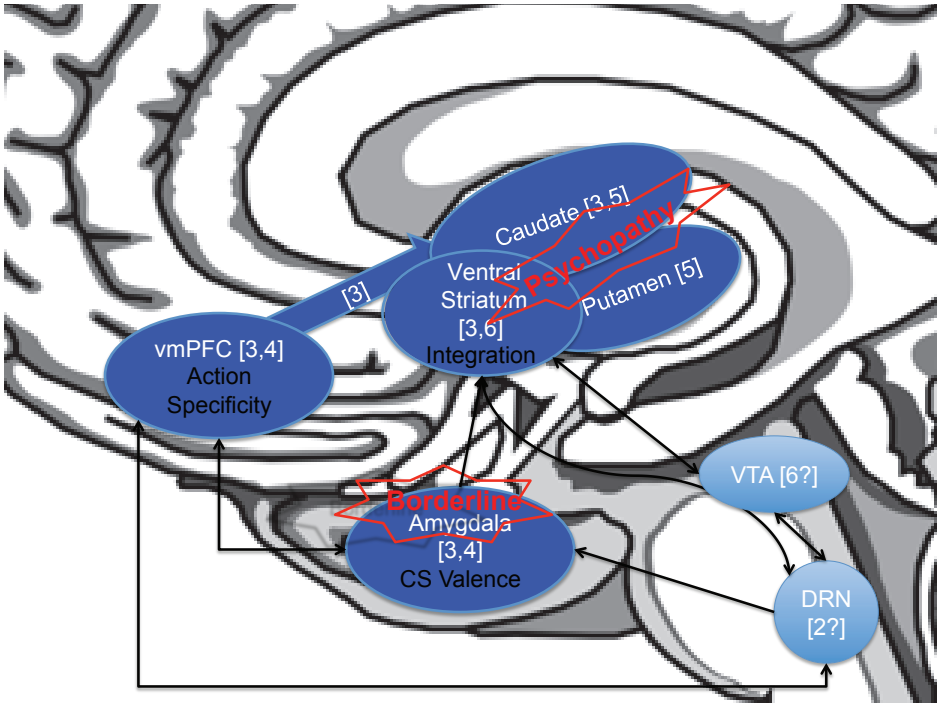


Figure 7.1 *– A schematic neural working model for aversive PIT derived from this thesis and extant literature. See text for description. Numbers between brackets refer to the specific chapters of this thesis. Light blue nodes and black connectors are not directly derived from this thesis. Cortico-striato-thalamico-cortical and striato-nigral-striatal loops are not displayed for clarity.*

substantia nigra pars compacta (Cardinal et al., 2002; Balleine and Doherty, 2009): In both the study by Talmi et al (2008) and in chapter 3 similar effects were found in the amygdala as in the ventral striatum. Moreover, functional disconnection of the basolateral complex of the amygdala and the nucleus accumbens shell abolished outcome-specific PIT in a study by Shiflett et al (Shiflett and Balleine, 2010). Extending this latter finding to aversive, outcome-general PIT would mean that disconnection of the central nucleus of the amygdala and nucleus accumbens core, would abolish the PIT effect (Corbit and Balleine, 2016). This should be tested in a future study.

In the model I propose that action specificity of the transfer effect is controlled by input from the vmPFC to the ventral striatum and head of the caudate nucleus. This hypothesis is based on findings that trial-by-trial instrumental actions were differentially represented in the vmPFC, depending on whether they were approach or withdrawal actions (chapter 3 and 4). Moreover, the interaction of this action-specific signal with the ventral striatum and head of the caudate nucleus was related to behavioural PIT effects in terms of button presses (chapter 3). This PIT-related functional connectivity is in accordance with anatomical studies indicating the connection between the vmPFC and ventral parts of the striatum as part of a set of parallel but interacting fronto-striatal loops (Haber, 2003; Haber and Rauch, 2010). These loops are suggested to integrate emotions, motivation, cognition, motor planning and execution of this planning (Haber, 2003). Furthermore, this proposal is in line with the functional role of the vmPFC in a wide range of cognitive functions relating to decision making, amongst others in context-dependent stimulus valuation (Rudorf and Hare, 2014) and instrumental (esp. goal-directed) behaviour (de Wit et al., 2009). The function of the vmPFC in goal-directed behaviour is thought to come about in interplay with the caudate nucleus (de Wit et al., 2012). For the vmPFC to adjust the motivational force of an aversive CS, I hypothesize that the vmPFC and/or its output to the striatum is modulated by output of the amygdala in reaction to CS display. Connectivity between the amygdala and vmPFC has been established anatomically as well as functionally (Kim et al., 2011; Myers-Schulz and Koenigs, 2011). Interaction between the amygdala and vmPFC has been proposed to enable organisms to react to biologically relevant predictive stimuli as well as adapt these reactions to situational circumstances (Kim et al., 2011). Thus, I propose that the vmPFC might be the critical structure for adjusting the motivational force that an aversive CS exerts on instrumental behaviour in a context specific manner. Future studies might assess whether modulation of vmPFC activity diminishes the PIT effect or only modulates action specificity or both. This is

especially interesting given the finding of Huys et al (2016), that loss of action specificity is related to major depression and ussuccessful recovery during treatment.

In summary, I hypothesize that the amygdala is involved in processing the affective aspects of the aversive Pavlovian cue. These affective/motivational aspects of the cue are translated to (action-specific) behavioural modulation via interaction with a fronto-striatal network consisting of the vmPFC, ventral and dorsal striatum.

7.4 SEROTONIN AND THE MOTIVATIONAL INFLUENCE OF AFFECT ON BEHAVIOR

In line with previous theories on the role of serotonin in decision-making (Cools et al., 2008; Dayan and Huys, 2008; Crockett et al., 2009; Dayan and Huys, 2009; Boureau and Dayan, 2011; Cools et al., 2011) data from chapter 2 suggests that serotonergic pathways are involved in the motivational control of instrumental behaviour by aversive affective cues. Indeed serotonin projections from the dorsal raphe nucleus are well known to modulate brain regions that are key for PIT, including the amygdala, prefrontal cortex (PFC) and striatum. Moreover, interactions between the vmPFC and dorsal raphe nucleus and serotonergic receptors in the medial PFC have been implicated in regulating approach-avoidance behaviours (Groenewegen and Uylings, 2000; Challis and Berton, 2015). Notably functional amygdala-vmPFC connectivity is modulated by serotonin and the serotonin transporter in emotion regulation (Hariri and Holmes, 2006; Fisher et al., 2011; Volman et al., 2013; Puglisi-Allegra and Andolina, 2015). This might be the pathway through which motivational forces and approach and withdrawal context distinctions are integrated (see previous section). In addition, the ventral striatum receives projections from both the amygdala and vmPFC and might as such interface with further action planning.

My finding that tryptophan depletion elicited aversive Pavlovian disinhibition concurs with recent theorizing about serotonin's role (Dayan and Huys, 2009; Boureau and Dayan, 2011; Cools et al., 2011; Faulkner and Deakin, 2014) but contrasts remarkably with a finding from another recent tryptophan depletion study (Hebart and Gläscher, 2015). As did our study, that study observed effects of serotonin depletion only in aversive, but not appetitive PIT. However, in their hands, serotonin depletion enhanced rather than reduced aversive PIT. Notably, in contrast to our study, the participants in the study of Hebart et al. did not exhibit any aversive inhibition under baseline. In fact the response rate under BAL during the

aversive CS was, if anything, higher (not statistically significant) than during the neutral cue. One possibility is that serotonin does not drive aversive behavioural inhibition per se, but only affects behavioural inhibition when there is a motivational drive to inhibit behavior to begin with. Indeed, in a recent study we have shown that reductions in serotonin lead to increased behavioural vigour only if there is a motivational drive to inhibit behaviour at baseline (Ouden et al., 2015). Another reason for the discrepancy with our results might be that their 'aversive' instrumental action (i.e. shooting) cannot readily be conceptualized as an aversive action. Future studies are needed to further these putative explanations to more concrete insights.

7.4.1 *Scope of these findings with respect to patients*

Serotonin plays a crucial role in several psychiatric disorders evidenced by many first line treatments of psychiatric disorders by serotonergic drugs. Our finding that serotonin underlies the coupling of aversive value and behavioural inhibition has therefore potential to add to our mechanistic insight in psychiatric disorders associated with serotonin dysfunction. For example, borderline personality disorder is related to aberrant reactions to aversive stimuli (Rosenthal et al., 2008). An activation instead of an inhibition of behaviour due to aberrant transfer of this aversive value to instrumental behaviour (and maybe also instrumental mental processing (Huys et al., 2012; Mendelsohn et al., 2014)) might explain their loss of control. Serotonergic dysfunction may also (partly) account for disinhibited behaviour in psychopathic criminals (cf. chapter 5). Within this group of patients I found that their psychopathic severity in terms of the PCL-R was positively associated with behavioural disinhibition. Moreover, psychopathy severity in terms of the PCL-R has been associated (parametrically) with reduced serotonergic functioning (Soderstrom et al., 2001; 2003). Taken together this might suggest that within a group of psychopathic criminals individual differences in behavioural disinhibition might be due to differences in serotonergic transmission. Future studies that either measure (e.g. collect CSF) or manipulate (e.g. by administering an SSRI) serotonergic transmission in these patients might provide more direct information about this link (cf Fanning et al., 2014).

7.5 PSYCHIATRIC SYNDROMES AND MOTIVATION OF BEHAVIOUR

Did my findings on borderline personality disorder, psychopathy and criminality in chapter 4, 5 and 6 improve our understanding of the neurocog-

nitive mechanisms underlying the affective dysregulation of behaviour in these patient groups? Do these findings have clinical implications?

The study in chapter 4 can be seen as a 'bench to bedside' contribution, where I used a laboratory 'bench' test (PIT paradigm) to assess clinically relevant symptom improvement 'at the bedside'. I did this in a particularly heterogeneous patient group with large variability in treatment response. Our mechanistic approach helped identify a predictor that was associated with individual differences in treatment success, namely PIT-related amygdala signal. Such predictors on the individual level are crucial for improving individual outcomes of treatment regimes (Jones et al., 2015; Heinz et al., 2016). The next question is whether the use of such predictors can enhance the efficiency of treatment programs. Would patients for which I predict that DBT will have a low impact on symptom severity benefit more from other treatment regimes like mentalization based treatment, transference focused treatment, schema focused therapy or good general management of borderline personality disorder? Which other tasks could be developed to address the hypothesized different mechanisms underlying the different psychotherapeutic regimes? I view the current study as a stepping-stone for addressing these questions.

In chapter 5 we showed that individual differences in psychopathy severity were accompanied by individual differences in the motivational influence of aversive affective cues. This links the cognitive mechanism of aversive PIT to psychopathic severity. Chapter 6 explicitly contributes to our understanding of the neurocognitive mechanisms that are involved in severe criminal behaviour. However the findings of chapter 6, in contrast to those of chapter 5, resulted from comparisons between groups. I was not able to leverage these findings to the individual level. Future studies might use more specific measures of criminal behaviour to assess whether the reward expectancy related connectivity between the ventral striatum and medial prefrontal cortex is related to individual differences in criminality instead of psychopathy severity (the primary measure I used).

These findings do not have a direct clinical consequence yet, but behavioural (like aversive PIT) and neurocognitive measures (like reward expectancy related functional connectivity) have the potential to be used in addition to questionnaires and chart reviews in clinical practice. These measures might be a valuable addition to the more conventional, partly static measures (relying on historical information that does not change) to monitor changes in a patient and add to risk-assessments. Before this can be done, in analogy to trials conducted with more conventional measures

(Camp et al., 2013), longitudinal trials that relate recidivism to these neurocognitive measures should be conducted.

7.6 LIMITATIONS AND FUTURE STUDIES

Limitations of my work relate to: (i) individual differences in PIT; (ii) paradigm and study design; (iii) sample size; and (iv) control groups. Below I will discuss these limitations.

7.6.1 *Individual differences in the motivational control of instrumental behavior by affective cues*

I observed substantial inter-individual differences in PIT effects within our samples. Understanding the source of these individual differences requires further work. I will raise a few hypotheses.

Individual differences in instrumental control

One source of individual differences in the motivational control of instrumental behaviour by affective cues might stem from differences in instrumental behaviour. Although I did not find correlations between instrumental accuracy and transfer effects, this certainly does not exclude that the kind of instrumental control that was employed – i.e. goal-directed or habitual – influences the magnitude of the observed transfer effects. Such an influence has been evidenced in rodents by Holland et al., where the strength of outcome-general PIT was positively associated with the amount of instrumental training and not affected by outcome devaluation (Holland, 2004). Moreover, Sebold et al. (2016) revealed that participants whose response behaviour evidenced model-free instrumental control showed a larger PIT effect that those acting in accordance with model-based instrumental control. I did not test whether the instrumental behaviour observed in our studies was goal-directed or habitual nor did I assess whether subjects were inclined to model-based and/or model-free behavioural control. Following Sebold and colleagues, adding quantitative assessments of instrumental control in terms of model-based/model-free behaviour to the PIT procedure might improve explanation of the observed individual differences.

Individual differences in Pavlovian conditioning

There are also individual differences in the type of Pavlovian learning that people tend to exhibit. One well-known difference is that between sign-

and goal-trackers. Some individuals will approach or look at the CS (sign-trackers), others will approach or look at the place where the outcome will be delivered (goal-trackers). This difference in conditioned response has been linked to the strength of PIT in animals and recently also in humans (Garofalo and di Pellegrino, 2015). The latter study showed that goal-trackers in contrast to sign-trackers did not show any PIT-effect.

Individual differences in demand characteristics

Test subjects in psychological experiments might not always follow the instructions as intended by the researcher. The concept of "demand characteristics" refers to the presence of artifacts in experimental data due to test subjects being aware of what the researcher is investigating, or anticipating to find (Orne, 1962). This awareness can have different consequences for the observed behaviors of participants. In our studies, participants might have explicitly thought through the meaning of the Pavlovian background cues and acted on their supposed meaning. They also might have noticed the PIT effect on their behaviour and might have acted against or in line with this effect. In the studies presented in this thesis, I did not question participants afterwards what they thought I was measuring and whether they consciously acted against or in line with our hypothesis. Thus I was not able to address such a confound. Data suggest that such confounds, as awareness of the effect, can be related to the measured PIT effect (Talmi et al., 2008). Demand characteristics might thus explain part of the data in human PIT experiments. However there is evidence that the human PIT effect also covaries with subtle psychophysiological responses to Pavlovian stimuli that are unlikely to be under control of such declarative 'strategies'. A study by Ly et al. (2014) shows that bodily freezing in reaction to an aversive stimulus predicts the transfer effect of this stimulus on instrumental behaviour. In future studies, I will measure such reliable psychophysiological responses, while also enquiring about declarative awareness and/or presenting the Pavlovian stimuli subliminally.

7.6.2 *Sample size and study design*

The studies presented in this thesis rely on sample sizes that have been common in neuroimaging literature. However, in the past few years, considerable debates have developed around the validity of current scientific practice (Button et al., 2013). To maximize the information that can be derived from the studies presented in this thesis, a future meta-analysis would

be helpful as well as replication studies. What is readily overlooked in the case of replication studies is that these generally require considerably more power than the original study (see Button et al., 2013 for details). In addition to enabling replication, a larger sample size would allow us (in terms of statistical power) to address the various individual differences mentioned in the previous section.

For the patient studies a larger sample size should also be complemented with additional control groups. To specify the results from the BPD study at least two additional groups would be necessary: (i) A waiting group receiving no treatment and a group of BPD patients receiving a different treatment modality to assess whether the predicted symptom changes are specific to the DBT treatment; (ii) Another patient group (e.g. patients with depression) to assess whether the predicted symptom changes are specific to BPD. With regard to the psychopathy studies, a major improvement would be to add a criminal sample that scores low on psychopathic traits. This would help, in addition to the questionnaires I used, to specify whether our results are specific for psychopathy or for overt criminality.

7.6.3 *Specifics of the behavioural paradigm*

The fMRI PIT-paradigm used in chapter 3, 4 and 5 addressed aversive PIT as well as action specificity of aversive PIT. It was not sensitive to appetitive PIT (in chapter 3, 4 and 5). Developing a new comprehensive fMRI PIT-task that is sensitive to both appetitive and aversive PIT would help to address valence specificity of our findings. However, it is not clear what made our paradigm insensitive to the motivational influence of the appetitive valence. It might be the case that subjects in supine position generally do not like to receive juice in their mouth and therefore the appetitive juice did not achieve appetitive properties. Informally, participants commented on the aversive juice being very aversive. It might be that the focus on the aversive juice, made the appetitive juice only appetitive in terms of not receiving the aversive juice (i.e. a sort of safety signal). This could just as well be the case for the neutral CS, thus diminishing the difference between the appetitive and neutral CS. I did find differences in subjective liking of the neutral and appetitive CS by means of the visual analogue scale measurements. It is however questionable whether liking (compared with wanting) would be the best indicator of the motivational power of the CS (Berridge et al., 2009; Pool et al., 2015). Several strategies might be taken to circumvent the ambiguity of the appetitive CS. One manner would be to use monetary wins and losses in the Pavlovian stage instead of juices as was done in chapter 2.

A further step forward would be to develop an animal paradigm that could test appetitive and aversive PIT in approach and withdrawal in one task. Then this model would allow us to test the hypothesized neural model directly in the form of lesion studies or studies employing transient, topographically and neurochemically targeted interventions.

7.7 CONCLUDING REMARK

With this thesis I hope to have contributed to the scientific enterprise to gather and translate basic scientific insights to clinical psychiatric practice, also known as 'translational psychiatry'. Much work remains to be done before psychiatrists will notice a difference in their consultation room and patients will notice a substantial difference in their lives due to the informative work of cognitive neuroscientists. I am convinced that doctors will change their procedures when information is available that shows that these neurocognitive findings can inform the prognosis of a patient. To get to this information and achieve an improvement in patient care, studies are needed with respectable sample sizes and critically with study-designs targeted at gathering information about neurocognitive mechanisms, treatment effects and most importantly their interaction.

7.8 REFERENCES

BALLEINE BW, Doherty JPOA (2009) Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. Neuropsychopharmacology 35:48–69.

BALLEINE BW, Ostlund SB (2007) Still at the Choice-Point: Action Selection and Initiation in Instrumental Conditioning. Annals of the New York Academy of Sciences 1104:147–171.

BERRIDGE KC, Robinson TE, Aldridge JW (2009) Dissecting components of reward: 'liking', 'wanting', and learning. Current Opinion in Pharmacology 9:65–73.

BOUREAU Y-L, Dayan P (2011) Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 36:74–97.

BUCKHOLTZ JW, Treadway MT, Cowan RL, Woodward ND, Benning SD, Li R, Ansari MS, Baldwin RM, Schwartzman AN, Shelby ES, Smith CE, Cole D, Kessler RM, Zald DH (2010) Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. Nature Neuroscience 13:419–421.

BUTTON KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14:365–376.

CAMP JP, Skeem JL, Barchard K, Lilienfeld SO, Poythress NG (2013) Psychopathic predators? Getting specific about the relation between psychopathy and violence. Journal of Consulting and Clinical Psychology 81:467–480.

CARDINAL RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience and Biobehavioral Reviews 26:321–352.

CARTONI E, Puglisi-Allegra S, Baldassarre G (2013) The three principles of action: a Pavlovian-instrumental transfer hypothesis. Front Behav Neurosci 7:153.

CHALLIS C, Berton O (2015) Top-Down Control of Serotonin Systems by the Prefrontal Cortex: A Path toward Restored Socioemotional Function in Depression. ACS Chem Neurosci 6:1040–1054.

COHEN-HATTON SR, Haddon JE, George DN, Honey RC (2013) Pavlovian-to-instrumental transfer: Paradoxical effects of the Pavlovian relationship explained. Journal of Experimental Psychology: Animal Behavior Processes 39:14–23.

COOLS R, Nakamura K, Daw ND (2011) Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology 36:98–113.

COOLS R, Roberts AC, Robbins TW (2008) Serotoninergic regulation of emotional and behavioural control processes. Trends in Cognitive Sciences 12:31–40.

CORBIT LH, Balleine BW (2016) Learning and Motivational Processes Contributing to Pavlovian-Instrumental Transfer and Their Neural Bases: Dopamine and Beyond. Curr Top Behav Neurosci 27:259–289.

CROCKETT MJ, Clark L, Robbins TW (2009) Reconciling the role of serotonin in behavioral inhibition and aversion: acute tryptophan depletion abolishes punishment-induced inhibition in humans. Journal of Neuroscience 29:11993–11999.

DAYAN P, Huys QJM (2008) Serotonin, Inhibition, and Negative Mood. PLoS Comput Biol 4:e4.

DAYAN P, Huys QJM (2009) Serotonin in affective control. Annu Rev Neurosci 32:95–126.

DAYAN P, Seymour B (2013) Values and actions in aversion. In: Neuroeconomics: Decision making and the brain (Glimcher PW, ed). Academic Press.

DE Wit S, Corlett PR, Aitken MR, Dickinson A, Fletcher PC (2009) Differential Engagement of the Ventromedial Prefrontal Cortex by Goal-Directed and Habitual Behavior toward Food Pictures in Humans. Journal of Neuroscience 29:11330–11338.

DE Wit S, Watson P, Harsay HA, Cohen MX, van de Vijver I, Ridderinkhof KR (2012) Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. Journal of Neuroscience 32:12066–12075.

DICKINSON A, Balleine B (2002) The role of learning in the operation of motivational systems. In: Steven's handbook of experimental psychology: Learning, motivation and emotion, 3rd ed. (Gallistel CR, ed), pp 497–534. New York: John Wiley & Sons.

DOLAN RJ, Dayan P (2013) Goals and Habits in the Brain. Neuron 80:312–325.

FANNING JR, Berman ME, Guillot CR, Marsic A, McCloskey MS (2014) Serotonin (5-HT) augmentation reduces provoked aggression associated with primary psychopathy traits. J Pers Disord 28:449–461.

FAULKNER P, Deakin JFW (2014) Neuroscience and Biobehavioral Reviews. Neuroscience and Biobehavioral Reviews 46:365–378.

FISHER PM, Price JC, Meltzer CC, Moses-Kolko EL, Becker C, Berga SL, Hariri AR (2011) Medial prefrontal cortex serotonin 1A and 2A receptor binding interacts to predict threat-related amygdala reactivity. biol mood anxiety disord 1:2.

GARBUSOW M et al. (2016) Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. Addiction Biology, 21(3), 719–731.

GAROFALO S, di Pellegrino G (2015) Individual differences in the influence of task-irrelevant Pavlovian cues on human behavior. Front Behav Neurosci 9.

GROENEWEGEN HJ, Uylings HBM (2000) The prefrontal cortex and the integration of sensory, limbic and autonomic information. 126:3–28.

HABER SN (2003) The primate basal ganglia: parallel and integrative networks. Journal of Chemical Neuroanatomy 26:317–330.

HABER SN, Rauch SL (2010) Neurocircuitry: a window into the networks underlying neuropsychiatric disease. Neuropsychopharmacology 35:1–3.

HARIRI AR, Holmes A (2006) Genetics of emotional regulation: the role of the serotonin transporter in neural function. Trends in Cognitive Sciences 10:182–191.

HEBART MN, Gläscher J (2015) Serotonin and dopamine differentially affect appetitive and aversive general Pavlovian-to-instrumental transfer. Psychopharmacologia, 232(2), 437–451. http://doi.org/10.1007/s00213-014-3682-3

HEINZ A, Schlagenhauf F, Beck A, Wackerhagen C (2016) Dimensional psychiatry: mental disorders as dysfunctions of basic learning mechanisms. J Neural Transm.

HOLLAND P (2004) Amygdala–frontal interactions and reward expectancy. Current Opinion in Neurobiology 14:148–155.

HUYS QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P (2011) Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding Rangel A, ed. PLoS Comput Biol 7:e1002028.

HUYS QJM, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP (2012) Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees Maloney LT, ed. PLoS Comput Biol 8:e1002410.

HUYS QJM, Gölzer M, Friedel E, Heinz A, Cools R, Dayan P, Dolan RJ (2016) The specificity of Pavlovian regulation is associated with recovery from depression. Psychol Med:1–9.

JONES KA, Menniti FS, Sivarao DV (2015) Translational psychiatry-light at the end of the tunnel. Annals of the New York Academy of Sciences 1344:1–11.

KIM MJ, Loucks RA, Palmer AL, Brown AC, Solomon KM, Marchante AN, Whalen PJ (2011) The structural and functional connectivity of the amygdala: From normal emotion to pathological anxiety. Behavioural Brain Research 223:403–410.

LINEHAN MM (1993) Cognitive-behavioral Treatment of Borderline Personality disorder. New York: NY Guilford Press.

LY V (2014) Individual differences in bodily freezing predict emotional biases in decision making. Frontiers in Behavioral Neuroscience, 8, 237.

MCDANNALD MA (2009) The behavioral and neural mechanisms underlying conditioned suppression: A model of human anxiety Holland PC, ed. Available at: http://search.proquest.com/docview/304916166.

MENDELSOHN A, Pine A, Schiller D (2014) Between Thoughts and Actions: Motivationally Salient Cues Invigorate Mental Action in the Human Brain. Neuron 81:207–217.

MOGENSON GJ, Jones DL, Yim CY (1980) From motivation to action: functional interface between the limbic system and the motor system. Progress in Neurobiology 14:69–97.

MYERS-SCHULZ B, Koenigs M (2011) Functional anatomy of ventromedial prefrontal cortex: implications for mood and anxiety disorders. Mol Psychiatry 17:132–141.

ORNE MT (1962) On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American psychologist.

OUDEN den HEM, Swart JC, Schmidt K, Fekkes D, Geurts DEM, Cools R (2015) Acute serotonin depletion releases motivated inhibition of response vigour. Psychopharmacologia 232:1303–1312.

POOL E, Brosch T, Delplanque S (2015) Stress increases cue-triggered "wanting" for sweet reward in humans. Journal of Experimental Psychology. Animal Learning and Cognition, 41(2), 128–136.

PREVOST C, Liljeholm M, Tyszka JM, O'Doherty JP (2012) Neural Correlates of Specific and General Pavlovian-to-Instrumental Transfer within Human Amygdalar Subregions: A High-Resolution fMRI Study. Journal of Neuroscience 32:8383–8390.

PUGLISI-ALLEGRA S, Andolina D (2015) Behavioural Brain Research. Behavioural Brain Research 277:58–67.

BLAIR RJ (2013) Psychopathy: cognitive and neural dysfunction. Dialogues Clin Neurosci 15:181–190.

RESCORLA RA, Solomon RL (1967) Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychol Rev 74:151–182.

ROSENTHAL MZ, Gratz KL, Kosson DS, Cheavens JS, Lejuez CW, Lynch TR (2008) Borderline personality disorder and emotional responding: A review of the research literature. Clinical Psychology Review 28:75–91.

RUDORF S, Hare TA (2014) Interactions between Dorsolateral and Ventromedial Prefrontal Cortex Underlie Context-Dependent Stimulus Valuation in Goal-Directed Choice. Journal of Neuroscience 34:15988–15996.

SEBOLD M, Schad DJ, Nebe S, Garbusow M (2016) Don't Think, Just Feel the Music: Individuals with Strong Pavlovian-to-instrumental Transfer Effects Rely Less on Model-based Reinforcement Learning. Journal of Cognitive Neuroscience, 28(7), 985–995.

SHIFLETT MW, Balleine BW (2010) At the limbic-motor interface: disconnection of basolateral amygdala from nucleus accumbens core and shell reveals dissociable components of incentive motivation. European Journal of Neuroscience 32:1735–1743.

SODERSTROM H, Blennow K, Manhem A, Forsman A (2001) CSF studies in violent offenders: I. 5-HIAA as a negative and HVA as a positive predictor of psychopathy. J Neural Transm 108:869–878.

SODERSTROM H, Blennow K, Sjodin AK, Forsman A (2003) New evidence for an association between the CSF HVA: 5-HIAA ratio and psychopathic traits. Journal of Neurology, Neurosurgery, and Psychiatry 74:918–921.

TALMI D, Seymour B, Dayan P, Dolan RJ (2008) Human Pavlovian Instrumental Transfer. Journal of Neuroscience 28:360–368.

VOLMAN I, Verhagen L, Ouden den HEM, Fernandez G, Rijpkema M, Franke B, Toni I, Roelofs K (2013) Reduced Serotonin Transporter Availability Decreases Prefrontal Control of the Amygdala. Journal of Neuroscience 33:8974–8979.

Part V

APPENDIX

8

NEDERLANDSE SAMENVATTING

Onze handelingen worden beïnvloed door de emotionele ("affectieve") staat waarin wij ons bevinden. In dit proefschrift onderzoek ik de neurocognitieve basis van deze beïnvloeding bij mensen met en zonder psychiatrische aandoeningen. Een voorbeeld van de invloed van affect op ons handelen: Stel je voor dat je net te horen hebt gekregen dat je werkgever je aan het eind van de dag wil spreken. Je hebt goede ervaringen met je werkgever en weet dat hij erg tevreden is. Je verwacht iets positiefs, wellicht wel een promotie. Door deze verwachting voel je je prettig, 'lichter' en dat voel je niet alleen, je ziet het ook terug in bijvoorbeeld je manier van lopen: rechtop, vlotter, grotere passen. Je moet naar je volgende afspraak. Daar zal je sneller zijn omdat de positieve verwachting, je positieve affectieve staat, blijkbaar invloed heeft op je handelen (in dit geval je lopen). Maar, Stel je nu voor dat je aan het eind van de dag bij je werkgever moet komen en dat eerdere gesprekken juist altijd zeer onprettig zijn verlopen. Je verwacht dat hij mogelijk ontslag aan gaat kondigen. Je kunt je voorstellen dat je dan niet zo licht en vlot naar de volgende afspraak loopt, maar meer voorover gebogen, sloffend, 'met je ziel onder de arm'. Ook hier heeft je affectieve staat invloed op je handelen. Het is dezelfde handeling, maar je voert hem anders uit. Dit doe je niet omdat je een ander doel hebt, je doel is om naar de volgende afspraak te lopen; je doet het anders omdat je affectieve staat anders is. Met andere woorden, onze affectieve staat controleert voor een deel hoe wij onze doelgerichte (of instrumentele) handelingen uitvoeren. Dit heeft tot gevolg, dat als wij willen begrijpen waarom wij de dingen doen zoals wij ze doen, we dan ook moeten begrijpen hoe onze affectieve staat ons instrumenteel handelen motiveert. Met dit proefschrift probeer ik aan dit begrip bij te dragen.

Het voorbeeld uit de vorige alinea is een alledaags voorbeeld en je zult weinig last ervaren van de vertraging op bijvoorbeeld het lopen. Misschien zullen we zelfs wel voordeel van dit effect hebben: mogelijk heb je, doordat je een en ander wat trager doet, meer tijd om na te denken over een goede strategie om het gesprek aan te gaan. De invloed van een affectieve staat op instrumenteel handelen kan voor veel mensen echter ook vervelende consequenties hebben. Patiënten met een psychiatrische aandoening kunnen hier behoorlijk last van hebben. Denk aan mensen met depressie (die

volledig stil kunnen vallen door negatieve informatie) of met angststoornissen (die in paniek kunnen raken door kleine dingen die hen aan wat vervelends doen denken). In dit proefschrift keek ik naar een groep patiënten met persoonlijkheidsstoornissen, waarbij het effect van hun affectieve staat op hun instrumenteel handelen ook belangrijk lijkt. Dit waren mensen met een persoonlijkheidsstoornissen die gepaard gaan met o.a. impulsieve en agressieve symptomen. Zo is het handelen van patiënten die lijden aan borderline persoonlijkheidsstoornis vaak erg gevoelig voor negatieve informatie. Daarentegen lijkt het handelen van patiënten met een psychopathische persoonlijkheid juist te ongevoelig voor negatieve informatie . Mensen met psychopathie lijken juist te gevoelig voor positieve affectieve informatie. Als we het gedrag van de mensen met deze aandoeningen willen begrijpen zullen we inzicht moeten krijgen in hoe hun affectieve staat hun instrumenteel handelen motiveert.

Onze emoties en ons gedrag komen niet uit het niets. Ze worden bepaald in een samenspel tussen wat er in de buitenwereld gebeurt, wat wij daarvan al dan niet opvangen met onze zintuigen, de reactie van onze hersenen daarop en het daaruit voortvloeiende gedrag waarmee wij de buitenwereld en onze zintuigen weer be\"invloeden. In dit proefschrift heb ik gekeken naar hoe affectieve informatie uit de buitenwereld via de hersenen ons instrumenteel gedrag beïnvloedt. Het doel van de studies in dit proefschrift is tweeledig: 1. het vergroten van ons begrip van hoe de hersenen ervoor zorgen dat affectieve informatie instrumenteel gedrag (de)motiveert in gezonde mensen, en om 2. dit begrip te gebruiken om affectieve dysregulatie van gedrag in patiënten met affectieve, impulsieve en agressieve symptomen beter te begrijpen. In hoofdstuk 2 en 3 komt doel 1 aan bod: ik heb gekeken naar hoe instrumenteel gedrag verandert door verschillende vormen van affectieve informatie aan te bieden aan gezonde mensen terwijl zij instrumenteel gedrag vertoonden. Dit deed ik door mensen een computertaak te laten uitvoeren, waarin Pavloviaans Instrumentele Transfer (PIT; zie box 1 hoofdstuk 1) gemeten werd. PIT is de interactie (“transfer”) tussen instrumenteel gedrag en affectieve, Pavloviaans informatie. Deze PIT taak bestond uit 3 delen: in het eerste deel leerden mensen instrumenteel gedrag aan: ze kregen verschillende paddenstoelen te zien waarbij ze konden kiezen of ze de paddenstoel wilden verzamelen of niet. Als ze een keuze maakten (wel of niet verzamelen), kregen ze meteen feedback of ze door deze keuze geld wonnen of verloren. Zo leerden ze met hun gedrag een doel te bereiken (geld te verdienen). In het tweede deel liet ik de mensen verschillende plaatjes zien. Deze plaatjes werden gevolgd door iets positiefs, neutraals of negatiefs, zodat de plaatjes een affectieve lading kregen. Vervol-

gens, in het derde deel van de taak, liet ik de mensen weer de taak met de paddenstoelen spelen, maar nu liet ik ook de plaatjes uit het tweede deel op de achtergrond zien. De interessante uitkomst van deze taak is hoe deze affectieve plaatjes het instrumentele gedrag (de keuzes om geld te verdienen) beïnvloeden. Over het algemeen zagen we dat negatieve plaatjes op de achtergrond mensen remden om de paddenstoelen actief te verzamelen, maar juist stimuleerden om paddenstoelen actief te ontwijken (hoofdstuk 3, 4 en 5). Dus actieve benadering werd geremd, maar actief ontwijken werd gestimuleerd door een negatieve affectieve prikkel.

In hoofdstuk 2 heb ik specifiek gekeken naar wat voor een effect vermindering van de neurotransmitter serotonine op deze transfer had. Serotonine is een molecuul dat in de hersenen gebruikt wordt om signalen door te geven tussen bepaalde hersencellen. Dit molecuul is betrokken bij verscheidene psychiatrische stoornissen en het meest bekend als aangrijpingspunt voor medicatie tegen depressie en angst. Serotonine is geassocieerd met remming van gedrag en verwerking van negatieve informatie. Op basis van eerder opgestelde theorie\"en over de rol van serotonine in het koppelen van negatieve informatie aan remming van gedrag, verwachtte ik dat vermindering van serotonine de koppeling tussen negatieve informatie en remming van instrumenteel gedrag zou verminderen. Ik vond inderdaad dat als serotonine verminderd werd in gezonde mensen, (een deel van) deze mensen dan minder geremd werden door negatieve affectieve informatie (in vergelijking met de conditie waarbij serotonine niet verminderd werd). Deze resultaten bieden proefondervindelijke ondersteuning voor het idee dat serotonine inderdaad belangrijk is voor de koppeling tussen negatief affect en gedragsmatige remming.

In hoofdstuk 3 heb ik vervolgens in gezonde mensen middels fMRI (hersenscans) gekeken naar welke hersengebieden betrokken zijn bij de invloed van affect op instrumenteel gedrag Ik vond dat het fMRI BOLD-signaal (een afgeleide van hersenactiviteit) in zowel de amygdala (o.a. betrokken bij emotieverwerking) als het ventrale striatum (o.a. betrokken bij anticipatie van affectieve waarde en actie) geassocieerd was met gedragsmatige remming door negatieve affectieve prikkels. Het was al wel bekend dat deze gebieden betrokken waren bij het motiverende effect van positieve affectieve prikkels op gedrag, maar dat ze ook betrokken waren bij dit specifieke effect van negatieve affectieve prikkels op gedrag was nieuw. Ik vond ook dat BOLD-signaal in de ventromediale prefrontale cortex (deel van de voorkwab van de hersenen) verschilde naargelang mensen in de PIT-computertaak de paddenstoelen actief benaderden of juist actief ontweken. Het samenspel ("functionele connectiviteit") tussen de ventromedi-

ale prefrontale cortex en het striatum bleek ook beïnvloed te worden door negatieve affectieve prikkels, en wel zo, dat hoe meer deze connectiviteit bij een persoon beïnvloed werd, hoe meer verandering in gedrag je bij deze persoon kon zien. Al met al geeft dit aan dat modulatie van fronto-striatale interacties door negatieve affectieve prikkels betrokken is bij motivatie van instrumenteel gedrag.

In het tweede deel van dit proefschrift onderzocht ik of we de methode en bevindingen van de vorige onderzoeken met gezonden mensen kunnen gebruiken om betekenisvolle informatie over patiënten en/of hun behandeling te achterhalen. Hiervoor onderzocht ik twee patiëntengroepen: mensen met psychopathie (PP) die misdaden hadden begaan en mensen met borderline persoonlijkheidsstoornis (BPS). Zowel PP als BPS hebben ernstige gevolgen voor patiënten, hun omgeving en de maatschappij. Beide stoornissen worden gekenmerkt door een verstoorde invloed van m.n. negatieve emoties op aangeleerd gedrag. Bij PP lijkt er te weinig invloed van negatieve emoties te zijn, bij BPS teveel. Tijdens de PIT-computertaak die we hierboven beschreven, hebben wij met fMRI net als in hoofdstuk 3 het BOLD-signaal van mensen met PP en BPS gemeten in de hersenen.

De studie naar patiënten met BPS (hoofdstuk 4) liet zien dat hersensignaal in de amygdala het verminderen van symptomen één jaar na de start van psychotherapie voorspelde: verminderde reactie van de amygdala vóór behandeling was gerelateerd aan een beter herstel één jaar na de start van behandeling. Dit laat zien dat herstel over een periode van tenminste één jaar voorspeld kan worden door fMRI hersensignaal in de amygdala dat gerelateerd is aan negatieve emotionele informatie. Deze studie maakt een veelbelovende start met het verhelderen van de (neurocognitieve) factoren waarmee we mogelijk succes van psychotherapie kunnen gaan voorspellen.

In de PP studie (hoofdstuk 5) liet de PIT-computertaak in criminelen met psychopathie zien dat de ernst van psychopathie gerelateerd was aan een verminderd remmende invloed van negatieve affectieve informatie op gedrag. We vonden dat bij de PP-groep de reactiviteit van de amygdala tijdens het zien van de plaatjes in het tweede deel van de computertaak gekoppeld was aan de ernst van de PP. Verder zagen we dat binnen de PP groep het hersensignaal in de nucleus caudatus gerelateerd was aan de gedragsmatige remming. Echter konden we de bevinding van het amygdala signaal en de bevinding in de caudatus niet betekenisvol aan elkaar koppelen. In de niet-criminele, gezonde controle groep vonden we geen relatie tussen gedragsmatige remming en BOLD-signaal in de nucleus caudatus. Een hypothese die uit deze bevindingen volgt is dat agressie (een

karakteristiek van PP) wellicht voortkomt uit verminderde remming van gedrag door negatieve affectieve informatie. Deze verminderde remming door emotionele informatie komt mogelijk niet door een verminderde emotionele reactie op zich, maar mogelijk door een verminderd effect van deze reactie op hersengebieden die betrokken zijn bij instrumenteel gedrag, zoals de nucleus caudatus.

Een derde studie die ik deed met patiënten (hoofdstuk 6) ging over beloningsverwachting bij mensen met PP en crimineel gedrag. Naast het vermoeden op ongevoeligheid voor straf, zijn er ook aanwijzingen gevonden dat bepaalde trekken (m.n. antisociaal en impulsieve trekken) van de psychopathische persoonlijkheid een sterke samenhang vertonen met beloningsverwachting enerzijds en met een versterkte reactie van het ventrale striatum op deze beloningsvewachting anderzijds. Deze bevindingen waren echter gebaseerd op bevindingen bij niet criminele mensen die trekken van PP lieten zien en niet bij mensen met PP en crimineel gedrag. Om dit gat op te vullen heb ik gekeken naar beloningsverwachting bij (1) niet criminele mensen die weinig antisociale, impulsieve trekken vertonen, (2) criminele mensen die deze trekken veel laten zien en (3) mensen met PP en crimineel gedrag uit de TBS kliniek. Ten eerste vonden we dat ten opzichte van groep (1) met weinig trekken, dat groepen (2) en (3) met veel trekken een verhoogde reactiviteit in het ventrale striatum laten zien bij beloningsverwachting. Ten tweede vonden wij dat groepen (2) en (3) op hun beurt weer sterk verschilden in connectiviteit tussen het ventrale striatum en de dorsomediale prefrontale cortex. Al met al suggereert dit dat criminelen met PP en gezonde mensen die hoog scoren op bepaalde trekken van PP niet verschillen in de primaire reactie in het ventrale striatum, maar door de functionele verbinding tussen dit gebied en de dorsomediale prefrontale cortex.

Samengevat tonen de studies dat:

- serotonine betrokken is bij de motivationele invloed van negatieve affectieve informatie op instrumenteel gedrag,
- de amygdala en frontrostriatale circuits betrokken zijn bij de motivationele controle van instrumenteel gedrag door negatieve, affectieve prikkels,
- verminderd hersensignaal in de amygdala bij PIT symptoomvermindering één jaar na start van psychotherapie voorspelt bij mensen met borderline persoonlijkheidsstoornis,
- de ernst van psychopathie binnen een groep mensen met crimineel gedrag samenhangt met hoe sterk hun gedrag geremd wordt door

negatieve affectieve prikkels en dat dit mogelijk samenhangt met de werking of aansturing van de nucleus caudatus, en ten slotte

- mensen met psychopathie en crimineel gedrag verschillen van niet-criminele controles met vergelijkbare impulsieve/antisociale trekken in neurale verwerking van beloningsverwachting in het ventrale striatum.

9

DANKWOORD | ACKNOWLEDGEMENTS

Er is zo ontzettend veel om dankbaar voor te zijn. Dat geldt ook zeker als ik terugkijk op de afgelopen jaren. Wat een rijkdom om in een fantastische (sociale) omgeving een gezin te stichten, te mogen specialiseren in de psychiatrie en om dit te combineren met een intrigerend promotietraject. Er zijn ontzettend veel mensen die ik hiervoor dank verschuldigd ben. Ik wil graag iedereen die heeft bijgedragen bedanken en een aantal hiervan hieronder nog wat extra bedanken.

Graag bedank ik de **patiënten** die mij de mogelijkheid geven mijn vak uit te oefenen. Zij zijn het die mij dag in dag uit inspireren met hun verhalen over hun strijd om hun leven vast te houden, op te pakken en/of te verbeteren. Dit doen zij te midden van woelige, heftige klachten: een (vaak onderschatte) fenomenale prestatie. Om dan ook nog mee te doen aan wetenschappelijk onderzoek: fantastisch! Hartelijk dank dan ook aan de proefpersonen die de studies in dit proefschrift mogelijk hebben gemaakt. Dit laatste geldt natuurlijk ook voor de **proefpersonen** die we indeelden in de controle groepen.

Veel dank gaat uit naar mijn (co-)promotoren team bestaande uit **Roshan, Hanneke en Robbert-Jan**: **Roshan**, wat goed dat ik bij jou in de groep gekomen ben en wat genereus dat je me daar wilde hebben. Je was voor mij een super supervisor. Je enthousiasme voor de neurowetenschap, je drive om zaken analytisch te doorgronden, je inzet om je studenten te laten groeien zijn tomeloos. En ja, deze zin is (ook) weer spreektaal, en ik werk niet echt efficiënt naar een punt toe, de lezer haakt af. . . , en je moet geen zinnnen met 'En' beginnen, etc. Je hebt mij een heleboel dingen op een heleboel verschillende fronten bijgebracht: ontzettend bedankt daarvoor!

Hanneke, je was een fantastische co-promotor van begin tot einde; een supervisor, collega en vriendin. Je hebt mij o.a. studies leren runnen en leren programmeren. We hebben veel gesproken over de wetenschap en misschien nog meer over zaken daaromheen en daarbuiten. Jij bent essentieel geweest voor mijn promotietraject en gelukkig blijven we elkaar ook zien nu dat ten einde is! Bedankt!

Robbert-Jan, bedankt voor je attente interesse, je kritisch klinische blik en het prettige persoonlijke contact tijdens de afgelopen jaren. Je enthousi-

asme voor translationeel onderzoek en de (forensische) psychiatrie werken aanstekelijk.

I have been surrounded by a tremendously varied and colourful group of people at the DCCN. The **cognitive control group** has been expanding ever since it existed and the last 7 years I was privileged to work together with **Rohan, Esther, Marieke, Martine, Hanneke, Monja, Lieneke, Mirjam, Joost, Guillaume, Bram, Ruth, Daan, Jeff, Ili, Sean, Katinka, Payam, Ruth, Desiree, Jennifer S, Jennifer C, Kirsten, Joris, Peter and Yvonne**. Thank you for the pleasant atmosphere during our meetings, labweekends and lunches.

Sean, my British friend, you are and were an asset to my personal and scientific life. I enjoyed the extremes of sharing rooms and flying together: from a 'bird alarm' in Brabant to several days with you in Stockholm where part of nightlife consisted of questions about amputation. Most of all, however, I enjoyed our daily walks and talks in which you set out your unique personal, philosophical view on life. Although these walks might be less frequent nowadays, I hope to have many more of those in the future.

Frank, sprankelend, activistisch, altijd in beweging, mooie ideeën, beter in kantoortafeltennis dan ik; wat een voorrecht om jou jaren als kamergenoot te hebben gehad en wat mooi dat je paranimf wil zijn: "hard goan!!". **Daphne** en **Jolien**, dank voor jullie gezelschap op onze kamer in het laatste jaar van mijn onderzoekstijd. **Marieke, Martine, Mieke, Hanneke, Joris** en **Peter**, jullie vormen voor mij de eerste labgroep waar ik in gezeten heb. Dank voor de leuke en inspirerende meetings, persoonlijke gesprekken, samenwerkingen en advies in de eerste jaren. **Kristen**, thank you for your enjoyable company a.o. while testing the juice set-up fort the ATD study. I remember vividly when you could not find the button to turn of the pumps while I got flooded with 'nasty juice' while lying supine in the dummy scanner.

Katinka en **Inge** bedankt voor de samenwerking bij de psychopathie studie. Wat een onderneming! Hierbij gaat extra dank uit naar **Katinka** die onmogelijk veel werk verzet heeft om de mensen uit de TBS met speciale vervoersdiensten op de juiste tijd op de goede plek te krijgen. En dat lukte niet alleen met mensen uit de TBS. Je wist te zorgen voor fantastisch leuke BBQs bij jou thuis met ook weer de juiste mensen met de juiste middelen op de juiste plek.

The methylphenidate group, **Payam, Sean, Lieneke, Monja, Hanneke, Jennifer S** and **Jennifer C.,** what a job: over 100 subjects on and off medication. Thank you so much for our fruitful and enjoyable collaboration!

Wat een van de meest opvallende aspecten van het Donders is, is dat het zo gesmeerd loopt. Dat komt voor een heel groot deel op het conto van de geweldige technische groep en administratie van het DCCN. Bedankt **Marek, Tildie, Erik, Mike, Sander, Uriel, Nicole, Arthur, Berend ☺, Sandra, Joost, Petra en alle anderen** voor jullie ondersteuning. Een speciale noot voor **Paul**, omdat je ook een speciale noot bent (en dan heb ik het niet alleen over je accent): dank voor je geduld en ondersteuning bij de vele uren die ik in jouw scannerruimte heb doorgebracht.

Een groot deel van mijn PhD traject werd doorkruist en afgewisseld met klinische bezigheden. De psychiatergroep van het RadboudUMC en Pro Persona is hier bijzonder ondersteunend in geweest. Speciale dank voor **Aart, Anne, Arnt, Indira, Jan-Willem, Joost, Monica, Philip** en **Sevket** (vanuit het RadboudUMC) en **Judith, Mark, Sharon, Els** en **Mario** (vanuit Pro Persona) die op inspirerende wijze bij mijn werkbegeleiding en mentoraat betrokken zijn geweest. Verder wil ik **Anne, Pauline** en **Thom** bedanken voor hun supervisies psychotherapie. Stuk voor stuk waren dat vormende en bijdragende ervaringen.

Amongst the vast variety of enriching, enduring and enlightening experiences which my PhD/psychiatry residency entailed, two stand out: one has been my visit to London of which I will carry warm memories and lots of other (clinical, psychotherapeutic, organisational) baggage with me. This would not have been possible without the friendly and knowledgeable guidance of **Steve Miller** and the personality disorder team at Touchstone (South London and Maudsley Trust). Thank you **Penny, Cat, Matt, Alice, Simon, Pedro, Trudy, Patricia and Eleanor**. De andere ervaring is mijn leertherapie bij **Hennie de Kroon**. Ik wil haar hartelijk danken voor de (dynamische) inzichten en haar voorbeeld als clinicus. Ik hoop ooit nog een keer de moed te hebben daar terug te keren.

Wat mij ontzettend gesteund heeft en waar ik ontzettend van genoten heb in de afgelopen jaren is het contact met ons eerstejaars AIOS groepje. Jarenlang veel gedeeld en nog steeds! **Hiske, Niels, Take, Inge** en **Karlijn**, fijn dat jullie er zijn!

Veel van de bovenstaande collega's zijn vrienden geworden. Vriendschappen zijn niet alleen prettig, verrijkend, vervullend, maar voor mij ook noodzakelijk om de veerkracht die het onderzoekswerk en klinische werk van je vraagt op een prettige manier te kunnen dragen. Ik ben dan ook erg blij met de voor mij speciale vriendschappen die ik (niet altijd optimaal) heb kunnen onderhouden met een aantal belangrijke mensen in mijn leven. Omdat tekst toch geen recht doet aan deze relaties, ook hier niet meer dan een paar woordjes. **De Lestrades**, jarenlang samengewoond, lief en leed van

het gezinsleven gedeeld. **Moos, Anna en Saartje**, **Lotte en Sander**, jullie zijn een bewonderenswaardig gezin. **Sander**, wat is het goed dat je in 2002 bij Quick bent gaan voetballen. **Lucas**, je fijnzinnigheid, **Wouter**, onze wandelingen en je onvermoeibare verbeterdrang, **Lodewijk**, gezamenlijkheid en nuchterheid, **Walter**, verkwikkend onderhoudend, **Renoud**, je gezelligheid. Voor discussie en gezelschap in de breedte, de heren van **A.H.G. Senex captiosus**: **Berend, Renoud, Matthijs, Igo, Wouter, Bart, Sander, Tom, Tom, Dennis, Maarten, Martijn** en **Marnix**. Een maffe vorm van ongein, onderlinge steun en serieuze gesprekken hoop ik nog jaren te blijven vinden bij de **Fellows, Fellies** en **Felliminis**: **Jelle, Paul, Roline, Karen, Roy, Marloes, Dick, Lobke, Tim** en **Sander**, wat een fantastische club. De directe buren van de **Frans Halsstraat** wil ik graag bedanken voor de dagdagelijkse (en avondse en nachtelijke) gezelligheid!

En dit alles had natuurlijk niets kunnen worden zonder de liefde, lessen en ondersteuning die ik meegekregen heb en krijg van mijn ouders, ooms en tantes. **Pa** en **Ma**, mijn dankbaarheid is door de reikwijdte van wat jullie voor mij betekenen niet uit te drukken, desalniettemin: bedankt! Dit proefschrift had er in ieder geval niet gelegen als jullie niet op verschillende fronten zo ondersteunend en behulpzaam waren geweest. Dat laatste geldt in surplus voor **Corry** en **Jan**. Ik was nog jaren bezig geweest als jullie niet ontelbare keren met alle liefde van de wereld waren bijgesprongen in ons gezinsleven. **Berend, Floor, Marcel, Mariëlle, Froukje** en **Gijs**, lieve broers en zussen en ooms en tantes, het is fijn om met jullie een familie te vormen waarin we er voor elkaar zijn als het nodig is. **Brigitte, Rik** en **Ruud** bedankt voor de gezelligheid op tientallen gezellige familiedates in de afgelopen jaren. **Dieter**, du bist ein besondere Großvater für unsere Familie. Wir haben schon Vieles erlebt zusammen, und ich hoffe, daßwir noch viele Jahren Weihnachten, Ostern, Ferien und noch viel mehr mit dir verbringen können.

Roos, Mirhte en **Jasmijn**. Het is zoals Mirthe op een ochtend in bed zei: "Papa, je moet het boek gewoon afmaken!". Wat vind ik jullie lief zeg en wat zijn jullie fantastische kinderen! Jullie verrijken mijn leven met kleur, fantasie en heel veel warmte. We blijven heel veel knuffelen!

Levensgeluk en Lenny waren synoniemen geweest, als jij, **Lenny,** niet zoveel meer was geweest dan dat. Wat vind ik je prachtig en wat houd ik toch ontzettend veel van je! Ik geniet van ons leven nu en kijk uit naar de komende decennia samen.

10

LIST OF PUBLICATIONS

IN THIS THESIS

Peer reviewed

GEURTS, **D. E. M.**, Huys, Q. J. M., Ouden, den, H. E. M., & Cools, R. (2013). Serotonin and Aversive Pavlovian Control of Instrumental Behavior in Humans. Journal of Neuroscience, 33(48), 18932–18939.

GEURTS, **D. E. M.**, Huys, Q. J. M., Ouden, den, H. E. M., & Cools, R. (2013). Aversive Pavlovian Control of Instrumental Behavior in Humans. Journal of Cognitive Neuroscience, 25(9), 1428–1441.

GEURTS, **D. E. M.**, Borries, von, K., Volman, I., Bulten, B. H., Cools, R., & Verkes, R. J. (2016). Neural connectivity during reward expectation dissociates psychopathic criminals from non-criminal individuals with high impulsive/antisocial psychopathic traits. Social Cognitive and Affective Neuroscience, 11(8), 1326–1334.

Submitted/in preparation

GEURTS **D.E.M.**, van den Heuvel T., Huys Q.J.M., Cools R., "Amygdala response predicts clinical symptom reduction in patients with borderline personality disorder", submitted

GEURTS **D.E.M.**, von Borries K.A., Huys Q.J.M., Bulten E., Verkes R.J., Cools R., "Psychopathy severity is associated with reduced aversive Pavlovian to instrumental transfer", in prep

NOT IN THIS THESIS

Peer reviewed

SCHAAF, van der, M. E., De Lange, F. P., Schmits, I. C., **Geurts, D.E.M.**, Roelofs, K, van der Meer, J.W.M., Toni, I., Knoop, H., (2016). Prefrontal structure varies as function of pain symptoms in patients with Chronic Fatigue Syndrome, Biological Psychiatrie, accepted

SCHAAF, M. E., Schmits, I. C., Roerink, M., **Geurts, D. E.**, Toni, I., Roelofs, K., van der Meer, J.W.M., Toni, I., Knoop, H., (2015). Investigating neural mechanisms of change of cognitive behavioural therapy for chronic fatigue syndrome: a randomized controlled trial. BMC Psychiatry, 1–13.

JANSSEN, L. K., Sescousse, G., Hashemi, M. M., Timmer, M. H. M., Huurne, ter, N. P., **Geurts, D. E. M.**, & Cools, R. (2015). Abnormal modulation of reward versus punishment learning by a dopamine D2-receptor antagonist in pathological gamblers. Psychopharmacologia.

OUDEN, den, H. E. M., Swart, J. C., Schmidt, K., Fekkes, D., **Geurts, D. E. M.,** & Cools, R. (2015). Acute serotonin depletion releases motivated inhibition of response vigour. Psychopharmacologia, 232(7), 1303–1312.

GARBUSOW, M., Schad, D. J., Sebold, M., Friedel, E., Bernhardt, N., Koch, S. P., Steinacher, B., Kathmann, N., **Geurts, D.E.M.**, Sommer, C., Müller, DK, Nebe, S, Paul, S., Wittchen, H.U., Zimmermann, U.S., Walter, H., Smolka, M.N., Sterzer, P., Rapp, M.A., Huys, Q.J.M., Schlagenhauf, F., and Heinz A., (2016). Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. Addiction Biology, 21(3), 719–731.

SCHAAF, van der, M. E., Zwiers, M. P., van Schouwenburg, M. R., **Geurts, D. E. M.,** Schellekens, A. F. A., Buitelaar, J. K., Verkes, R.J., Cools, R., (2013). Dopaminergic drug effects during reversal learning depend on anatomical connections between the orbitofrontal cortex and the amygdala. Frontiers in Neuroscience, 7, 142.

SCHOUWENBURG, van, M. R., Zwiers, M. P., van der Schaaf, M. E., **Geurts, D. E. M.,** Schellekens, A. F. A., Buitelaar, J. K., Verkes, R.J., Cools, R., (2013). Anatomical connection strength predicts dopaminergic drug effects on fronto-striatal function. Psychopharmacologia, 227(3), 521–531.

HOLSTEIN, van, M., Aarts, E., van der Schaaf, M. E., **Geurts, D. E. M.**, Verkes, R. J., Franke, B., et al. (2011). Human cognitive flexibility depends on dopamine D2 receptor signaling. Psychopharmacologia, 218(3), 567–578.

SCHAAF, van der, M. E., van Schouwenburg, M. R., **Geurts, D. E. M.**, Schellekens, A. F. A., Buitelaar, J. K., Verkes, R. J., & Cools, R. (2014). Establishing the dopamine dependency of human striatal signals during reward and punishment reversal learning. Cerebral Cortex, 24(3), 633–642.

GEURTS, **D. E. M**., van der Velden, W. J. F. M., Hebeda, K. M., & Raemaekers, J. M. M. (2008). Richter's syndrome developing in a patient with adult onset Still's disease. Annals of Hematology, 88(1), 81–84.

THEELEN, T., Meulendijks, C. F. M., **Geurts, D. E. M.**, van Leeuwen, A., Voet, N. B. M., & Deutman, A. F. (2004). Impact factors on intraocular pressure measurements in healthy subjects. The British Journal of Ophthalmology, 88(12), 1510–1511.

Submitted

SWART, C. J., Froböse, M. I., Cook, J. L., **Geurts, D. E. M.,** Frank, M. J., Cools, R., & den Ouden, H. E. M., Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action, submitted.

OTHER PUBLICATION (NON PEER REVIEW)

COOLS, R., **Geurts, D. E. M.,** & van Holstein, M. (2012). Onderzoek naar gedragscontrole op het Donders Instituut . In M. Peeters, W. Meijer, & R. Verhoeff (Eds.), Wetenschappelijke doorbraken de klas in! (pp. 99–143). Nijmegen.

GEURTS, **D. E. M.** (2012). Kunnen robots zorgen? Of beter... kun je zorgen zonder mens te zijn?, Tijschrift voor Gezondheidszorg en Ethiek, 12(2), 57–58.

11

CURRICULUM VITAE

Dirk Geurts was born on the 9th of December 1983 in Doetinchem, The Netherlands. After he obtained his high school diploma at the Isala College (now Almende College) Silvolde in 2000, he studied Chemical Technology at Twente University. After obtaining his propedeutic diploma in 2001 he switched to study Medicine at the RadboudUMC and Philosophy at the Radboud University Nijmegen. He obtained his Bachelor in philosophy in 2005 with a special interest for philosophy of science and neo-Darwinian evolution mechanisms. In 2007 he obtained his Master's degree in philosophy under supervision of Prof. Dr. M. Slors with a thesis on the individuation of mental states. In the same year he received his "doctoraal" degree (equivalent to Master's title) in Medicine. After two years of clinical internships, a.o. three months in Hannover, Germany, Dirk obtained his Medical Degree. In 2008 he started a specialist training in psychiatry at the RadboudUMC and in 2009 his PhD project in the Cognitive Control group of Prof. Dr. Roshan Cools at the Donders Institute for Brain, Cognition and Behaviour. In 2010 he obtained an AGIKO ZON-MW grant to proceed with his combined trajectory as psychiatrist in training and PhD student. After a clinical residency with the South London and Maudsley Trust, London, United Kingdom, in 2016 he finished his clinical training. In 2017 he received the Koningsheide award for his contribution to the scientific field of forensic psychiatry based on the research presented in chapter 6. At the moment of publication of his thesis he works as a consultant psychiatrist and clinical lead in a crisis resolution home treatment team of Pro Persona in Tiel and is conducting research at the RadboudUMC and Donders Institute. Currently he is working on the effects of catecholamines (methylphenidate intervention) on the motivational control of instrumental behavior in the Cognitive Control group and assessing neurocognitive mediators of mindfulness based cognitive therapy in an RCT involving patients with ADHD in colaboration with Prof. Dr. Anne Speckens. Dirk is currently living in Nijmegen together with his wife Lenny Geurts-van Bon and his three daughters Roos, Mirthe and Jasmijn.

12

DONDERS GRADUATE SCHOOL FOR COGNITIVE NEUROSCIENCE

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit:

http://www.ru.nl/donders/graduate-school/donders-graduate/