Decoding reward signals in the brain: Functional neural correlates of individual reward valuation and reward learning

Justus-Liebig-Universität Gießen

Fachbereich 06

Psychologie und Sportwissenschaften

Otto-Behaghel-Straße 10F

35394 Gießen, Deutschland

13.09.2022

Synopsis zur kumulativen Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

Vorgelegt von

Sanja Klein

Dekanin: Prof. Dr. Christiane Hermann (Gießen, Psychologie)

Betreuer*innen:

Prof. Dr. Rudolf Stark (Gießen, Psychologie)

Prof. Dr. Jana Strahler (Freiburg, Sportwissenschaft)

Acknowledgements

I would like to thank…

…my supervisor Prof. Rudolf Stark for giving me the freedom to explore my own ideas. I am grateful for all the opportunities his lab offered me.

…Prof. Jana Strahler for being my supportive second supervisor and Prof. Tim Klucken for assessing this thesis.

…Onno Kruse for his scientific support, encouragement, general mentorship and dealing with my many many questions over the years.

…my colleagues at the lab for their helpful advice, in particular Isabell Tapia León for her experimental design advice, and Axel Schäfer, Bertram Walter and Carlo Blecker for their enduring stats and tech support. Above all, I appreciated the many coffee breaks, fun evenings, unhinged christmas parties and conference shenanigans with my fellow PhD students.

…my family for their unrelenting support, making this thesis possible in the first place and for always being interested in my research no matter how weird it sounded to them.

…my friends for every fun evening of cooking, board games, cocktails, jam sessions, knitting and venting in solidarity.

…Lena for sticking with me through the Good, the Bad and the Ugly of this PhD. Four years can feel like eternity, sometimes I shook my head and it was empty but you helped me come out on the other side. Thank you for the perpetual proofreading, moral support and every little thing you did to make this no waste of an education.

Contents

I Synopsis 2

1. Overview 2

2. Introduction 3

Rewards & Reward Value 3

Dopamine & Reward Prediction Error 4

Learning to predict reward 5

Neural correlates of reward prediction 7

3. Summary of Published, Peer-Reviewed Articles 12

Study I: Subjective reward value of visual sexual stimuli is coded in human striatum and orbitofrontal cortex (Klein et al., 2020) 12

Study II: Similarity between neural activity during appetitive conditioning and neural signature of fear conditioning (Klein et al 2022/under review) 15

4. Discussion 17

5. References 21

II Publications 35

Publication 1: Subjective reward value of visual sexual stimuli is coded in human striatum and orbitofrontal cortex 35

Publication 2: Cross-paradigm integration shows a common neural basis for aversive and appetitive conditioning 44

III Appendix 57

List of all publications 57

Statement regarding good scientific practice 59

I Synopsis

1. Overview

In this dissertation, I cover two research questions related to value and salience aspects of the neural representation of reward: First, how the individual value of immediate rewards is represented in the brain and second, which parts of learned reward expectations in the brain may be salience-based and thus independent from value. I will present an introduction with a theoretical overview of the concepts reward, motivational value and salience, reward learning and the current state of evidence on their neural correlates. Thereafter, I present the two studies in which I address the research questions concerning value of immediate rewards (visual sexual stimuli; Klein et al., 2020) and learned salience by assessing similarities between learned reward and fear, i.e. appetitive and aversive conditioning (Klein et al., 2022).

My other publications not directly part of this dissertation mainly relate to different aspects of neural processing of sexual rewards (Markert et al., 2021; Stark et al., 2019; Stark et al., 2022; van 't Hof et al., 2021; Klein et al., in press), appetitive conditioning processes (Klucken et al., 2019; Kruse et al., 2020) as well as more general emotion processing in the brain (Klein et al., 2019). We found the expectation as well as immediate viewing of sexual pictures and videos robustly associated with activation in reward-related brain regions in mixed-gender (Stark et al., 2019) as well as men-only samples (Markert et al., 2021). Under acute psychosocial stress, we observed increased brain activation during expecting and viewing sexual videos (Stark et al., 2022). Using a machine learning approach, we also developed a neural signature related to immediate viewing of sexual stimuli, which included many of the regions identified in our earlier research (van 't Hof et al., 2021). Concerning clinical implications, I reviewed research on how these and other aspects of sexual reward processing factor into problematic pornography use (Klein et al., in press). In the appetitive conditioning studies, we used monetary instead of sexual stimuli as rewards. Using these paradigms we could show how neuroticism is associated with decreased conditioned neural responses (Klucken et al., 2019) and which regions are robustly involved in appetitive extinction (Kruse et al., 2020). Finally, in my first publication, I found heightened neural reactivity to positive and negative affective pictures associated with long-term cumulative testosterone levels in

men, pointing to increased emotional reactivity regardless of emotional valence (Klein et al., 2019).

For the two main research questions, I present an overview of the study methodology and the results within their context. In the concluding discussion, I interpret and integrate the results, discuss possible clinical implications as well as an outlook for future directions.

2. Introduction

Rewards & Reward Value

Rewards are stimuli that have the potential to induce positive emotion, approach behavior and learning. Primary rewards are the nutrients and liquids needed for homoeostasis as well as activities needed to mate, have children and care for children. Nonprimary rewards (e.g. money, luxury goods, gourmet food, art etc.) can have sensory properties that we experience as pleasurable and are ultimately also related to obtaining homeostatic goals (Schultz, 2015). For example, the associations between money, buying food and eating that food has to be learned first before money is experienced as rewarding. Thus, the effects of nonprimary rewards depend on the individual learning history of each person. When we encounter rewards, we assign them a subjective value. According to Schultz (2015), the subjective reward value is not defined only by physical reward properties but is represented in subjective preferences and choices. It is determined from characteristics of the situation (e.g. effort needed to obtain the reward), person (e.g. preferences, satiety) as well as sensory and physical properties of the rewards (e.g. visual, tactile, magnitude, delay). Therefore, the subjective reward value is internal and varies between different persons and situations (Schultz, 2015). In common theories (Robinson & Berridge, 1993; Zhang et al., 2009), the subjective reward value is separated into a motivational component, i.e. how much a reward is 'wanted', and a hedonic component, i.e. how much a reward is 'liked'. However, debates on how value is computed in the brain and to what degree these components are even separable, are still ongoing. Furthermore, motivational and hedonic components of value are both positively valenced and a main point of contention is how an unsigned motivational signal i.e. salience can fit into this system (Bromberg-Martin et al., 2010). One likely very important

mechanism for signed (value) and unsigned (salience) motivational signals is the behavior of dopaminergic neurons throughout the brain.

Dopamine & Reward Prediction Error

Reward value can be extracted from a multitude of different information sources including physical (e.g. glucose, temperature, intestinal filling detectors) and sensory receptors (e.g. visual, auditive, gustatory, tactile) but also past learning experiences related to the reward (Marchner & Preuschhof, 2018). It is believed that this integration of information sources and ultimate extraction of value mainly happens in the dopaminergic system (Arias-Carrión et al., 2010). In animal studies, the receipt of a strong, unexpected reward has been shown to increase phasic firing of dopamine neurons (Schultz, 1998) while rewards that are fully predictable produce only little response (Montague et al., 1996; Schultz et al., 1997). Therefore, it is believed, that the activity change in dopamine neurons represents an error in the prediction of the value of an immediate or future reward – the reward prediction error (RPE). Strong phasic dopamine excitement indicates that a reward is better than expected, phasically inhibited dopamine neurons indicate a reward worse than expected (Schultz et al., 1997). Thus, the subjective reward value is equivalent to the RPE when the reward is presented immediately. The RPE for future rewards is crucial as a learning signal (Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972). In Temporal Difference Learning (Sutton & Barto, 1987) the prediction error is defined as the difference between the expected value of all future reward at the current point in time and the expected value of all future reward at a later, succeeding point in time, discounted by how much time lies between these points. Learning is strongest with a large initial RPE and it gets smaller as the prediction gets better. When the RPE equals zero, no (more) learning takes place. Finally, it is important to note that these learning models are not restricted to rewards. Learning can also take place in a similar way when the outcomes are only worse than expected – a type of aversive prediction error for example for electric shocks, loss of money or negative feedback (Seymour et al., 2004). Aversive prediction errors reflect that a stimulus is relevant (increased motivational salience), similarly to RPEs but also that it is unpleasant (decreased motivational value). It is not clear as of yet, whether all prediction errors are coded in similar ways and how the differentiated valence information would be contained, especially in humans. Animal evidence suggests that while many dopamine neurons respond to increased reward only (motivational value signal),

others respond to both increased and decreased reward (unsigned motivational salience signal; Bromberg-Martin et al., 2010, see figure 1). To find out, how motivational value is represented in the human brain on an individual basis, I examined the processing of sexual rewards, which are highly biologically relevant but also highly dependent on individual preferences. In my second study, I aimed to assess how much of the brain activation pattern in appetitive learning is associated with motivational salience by quantifying the commonalities with an aversive learning pattern. Therefore, I first briefly introduce the used learning paradigms before elaborating on the neural correlates of reward.

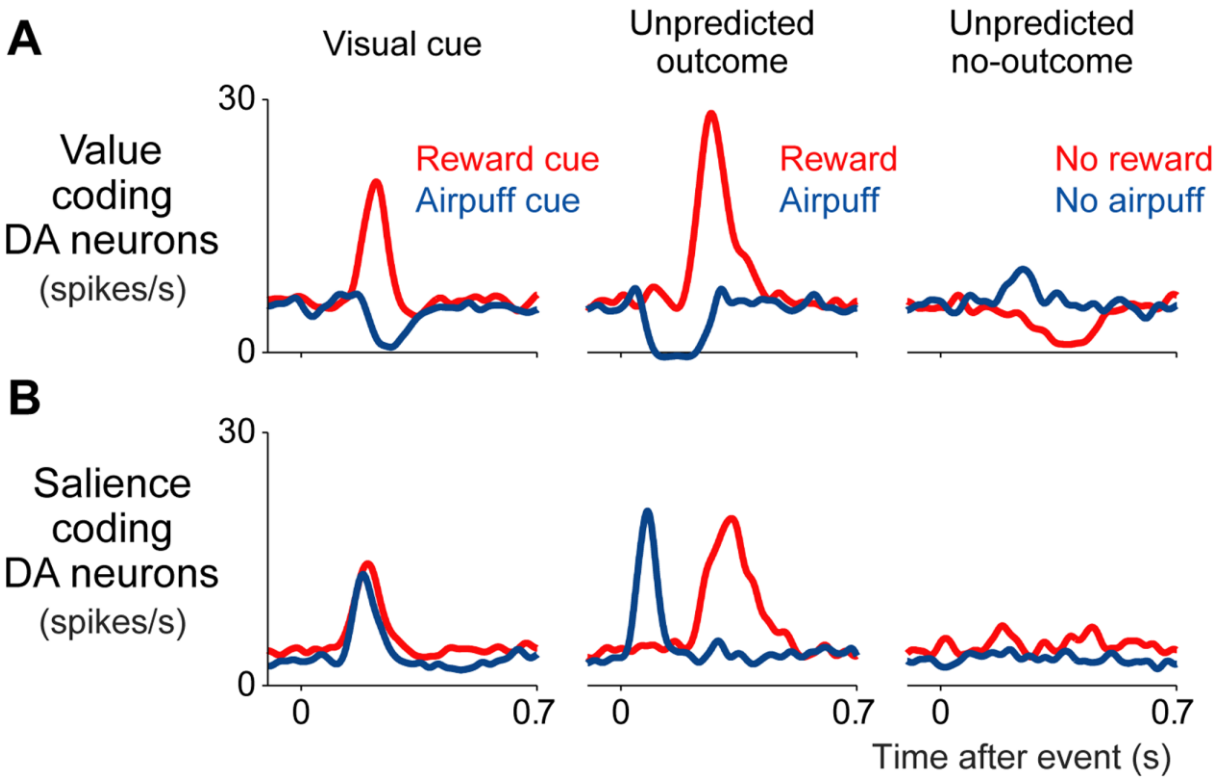

Figure 1: Distinct groups of midbrain dopamine neurons coding motivational value and salience in monkeys. (A) Motivational value neurons are excited by stimuli predicting appetitive outcome (juice) and inhibited by stimuli predicting aversive outcome (airpuff). (B) Motivational salience neurons are excited by both stimuli predicting appetitive and aversive outcomes. Figure adapted from Bromberg-Martin et al., 2010 with original data from Matsumoto & Hikosaka, 2009.

Learning to predict reward

In human fMRI studies, many different paradigms can be used to examine reward learning processes, depending on where the focus lies. Appetitive conditioning is a form of learning through association. It describes the process whereby an initially neutral stimulus (NS)

becomes a conditioned stimulus (CS+) after repeated pairing with a rewarding (appetitive) unconditioned stimulus (UCS, e.g. money). In differential conditioning paradigms, a second stimulus (CS-) is never paired with the UCS. The UCS triggers an unconditioned response (UCR, e.g. heightened skin conductance, increased arousal or valence ratings, increased approach behavior and activation in reward-related brain areas). After repeated CS+/UCS pairing, the CS+ then triggers a conditioned response (CR), which is similar to the UCR (Baeyens et al., 1990; Mackintosh, 1975).

An important characteristic of appetitive conditioning tasks is whether they are purely passive and require no action from the participant (i.e. classical conditioning, first described by Pavlov, 1927) or if the participant needs to perform actions which are then rewarded (instrumental conditioning, first described by Skinner, 1937). In classical appetitive conditioning tasks, CS+ and UCS become associated merely through observing the repeated pairing as described above. In instrumental appetitive conditioning tasks, participants are rewarded when they perform a specific behavior (such as pressing a lever or button). This behavior is reinforced and becomes more likely to be shown again. To examine these processes combined in human fMRI studies, an adapted monetary incentive delay task (MID task; Haber & Knutson, 2010) can be used. In the MID task, a fast reaction to a target stimulus is only rewarded with money after a CS+ has been presented but not after a CS- has been presented. In the original version, participants are instructed which stimuli are CS+ and CS- before performing the task. In the adapted version, the instruction is omitted, thus learning which stimuli are CS+ and CS- takes place mainly during the task (Kruse et al., 2017; Kruse et al., 2020).

It is vital to gain knowledge on reward learning processes since they are an important mechanism in many psychological disorders, most notably the development and maintenance of addictions (Martin-Soelch et al., 2007). For example, the motivational value and salience of rewarding stimuli (e.g. drugs, food) and behaviors (e.g. gaming, gambling, exercise, viewing pornography) can influence learning in the early and later stages of addiction development. The incentive sensitization theory of addiction postulates that the repeated use of substances sensitizes the reward circuitry to cues associated with these substances, over time attributing greatly increased motivation to them (Robinson & Berridge, 1993). Thus, how motivational

value and salience is assigned to rewarding stimuli or behaviors and changes over time is crucial to better understand all kinds of addictive disorders.

Neural correlates of reward prediction

In the human brain, dopaminergic neurons are mainly concentrated in the mesencephalon (in the substantia nigra and ventral tegmental area, VTA), with projections to the striatal nucleus accumbens (NAcc), caudate nucleus, putamen and the prefrontal cortex (dorsal, ventral and orbitofrontal PFC). These regions with their interconnections are commonly referred to as the mesocorticolimbic dopamine system (see figure 1; Arias-Carrión et al., 2010). The amygdala and the cerebellum are closely interconnected with many of these dopaminergic regions (Carta et al., 2019; Kita & Kitai, 1990). Reward processing in these areas can be studied using functional magnetic resonance imaging (fMRI) since recording single neurons is not feasible in humans. Other regions, which I do not focus on in detail here, but are likely important for appetitive prediction include the dorsal and ventral anterior cingulate cortex (dACC, vACC), the thalamus and the insula (Chase et al., 2015; Martin-Soelch et al., 2007).

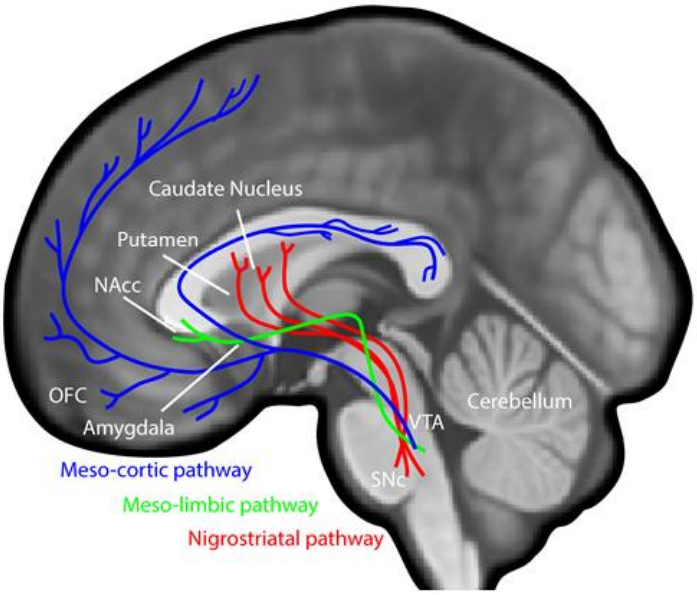

Figure 2: The mesocorticolimbic dopamine system. Adapted from Arias-Carrión et al., 2010)

The striatal subregions are functional correlates of appetitive prediction errors concerning immediate and future rewards (Bartra et al., 2013; Kühn & Gallinat, 2012; Peters & Büchel, 2010), specifically, the nucleus accumbens (NAcc), as part of the ventral striatum, and the caudate nucleus as part of the dorsal striatum (Haber & Knutson, 2010). The NAcc responds to immediately rewarding stimuli as well as stimuli that signal an obtainable future reward or an avoidable future loss, i.e. stimuli with high motivational value as well as motivational

salience (Oldham et al., 2018; Wilson et al., 2018). Activation in the NAcc has been shown to increase with increasing magnitude of monetary reward (Knutson et al., 2001) and to correspond to individual ratings of rewards (Cloutier et al., 2008; Lebreton et al., 2009; Rolls et al., 2008). These studies support the NAcc coding mainly motivational value and also salience components of appetitive prediction errors concerning immediate and future outcomes. In terms of learning, it is therefore very important in instrumental conditioning, as it is believed to code the motivational properties associated with the CS+ (Delgado, 2007; Garrison et al., 2013). One explanation for this is that different neuronal populations in the ventral tegmental areal affect different anatomical subdivisions of the NAcc, one conveying a general motivational salience signal (NAcc core) and one a signed motivational value signal (NAcc shell; Bromberg-Martin et al., 2010). Animal studies suggest that the medial NAcc shell functions as an 'affective keyboard', generating positively valenced emotions at its anterior end and increasingly negatively valenced emotions towards the posterior end (Berridge & Kringelbach, 2015). The caudate nucleus is involved in preparing and taking actions to obtain a reward (Balleine & O'Doherty, 2010; Everitt & Robbins, 2013). Caudate activity is increased when participants actively choose a monetary reward that is higher in magnitude or less delayed in its delivery (Le et al., 2020; Luo et al., 2009; Pine et al., 2009). Taken together, the evidence suggests that the caudate nucleus encodes appetitive prediction errors for a goal-directed action based on the expected outcome. Because of this involvement in outcome-driven action selection, it is likely essential for instrumental learning (Balleine & O'Doherty, 2010). The putamen, which is also part of the dorsal striatum, is involved in forming stimulus-response-associations that are independent from changing outcome values (Everitt & Robbins, 2013). This indicates that the region does not encode appetitive prediction errors in a similar way to the other striatal regions. The putamen is likely still crucial for learning, specifically the development of habits out of formerly outcome-directed actions (Tricomi et al., 2009).

The orbitofrontal cortex (OFC) is also implicated in reward value processing (Kühn & Gallinat, 2012; Peters & Büchel, 2010). Research has shown activation in the OFC to scale with individual pleasantness reports for many different rewards (Cloutier et al., 2008; Hare et al., 2009; Lebreton et al., 2009; Plassmann et al., 2008; Rolls & McCabe, 2007; Royet et al., 2003), with the expected magnitude of rewards (Kim et al., 2011; Sescousse et al., 2010) as well as expected reward and punishment (Metereau & Dreher, 2015). While the medial OFC shows

appetitive prediction error signals across reward types according to a 'common currency' theory (Bartra et al., 2013; Clithero & Rangel, 2014), more lateral and lateral/central parts have shown differentiated signals, segregated by reward type (Klein-Flügge et al., 2013; Sescousse et al., 2010). Taken together, the OFC seems to encode a more abstract and hedonic 'liking' component of the predicted rewards, but not necessarily the motivational component contained in the striatal response. It has also been found to continuously update this prediction (Howard & Kahnt, 2021) and encode potential future rewards (Nassar et al., 2019; Wimmer & Büchel, 2019), making it essential for both instrumental and classical learning processes.

The amygdala, located below the striatum, includes the basolateral amygdala (BLA) and the central nucleus of the amygdala (CeN; Janak & Tye, 2015). The BLA has been found involved with associating the positive UCS value with appCS+ and in turn with encoding avCS-UCS associations (Everitt & Robbins, 2013; LaBar et al., 1998). The CeN has been suggested by lesion studies to have a role in avCS-UR associations, thereby promoting expression of the CR (LeDoux et al., 1988). This view has been extended by more recent animal work. This work found distinct and overlapping neuronal populations in the BLA associated with assigning emotional value to aversive as well as appetitive UCS and the CS+ that predict them (Janak & Tye, 2015). The CeN is now considered critical for CS-CR associations in both aversive as well as appetitive conditioning (Ciocchi et al., 2010; Warlow et al., 2017).

The cerebellum has been in focus of human conditioning research for a shorter time although its role in predicting outcomes of motor behavior has been known for a while (for a recent review see Popa & Ebner, 2018). Some studies have reported on the role of the cerebellum in aversive (Ernst et al., 2019; Lange et al., 2015) as well as appetitive prediction (Heffley & Hull, 2019; Klucken et al., 2013; Lam et al., 2013) but the region has been often overlooked in conditioning literature (Tovote et al., 2015).

Summary of theoretical overview and research questions

With my research projects, I aimed to get a better understanding of both motivational value and salience in the human brain. We know from animal studies that many different brain regions are involved in extracting value and salience information from reward. Human

neuroimaging studies have corroborated these accounts for money as a universal secondary reward but also more primary rewards like food and more abstract ones like music (Kühn & Gallinat, 2012). So, while a lot is known already on motivational value in the human brain for these rewards, the picture is less clear for other types of reward.

Visual sexual stimuli (VSS) are a type of biologically relevant reward that directly elicits positive outcomes (sexual arousal), but this reaction also depends on previously learned associations between VSS and sexual activity. These factors contribute to the predicted reward value of VSS which is thus highly dependent on individual preferences. This makes VSS ideal to use for examining reward value signals at an individual level. Furthermore, problems related to excessive VSS use such as Problematic Pornography Use (PPU) are wide-spread and researchers assume that individual preferences play a larger role here than in other behavioral addictions (Brand et al., 2016; Brand et al., 2019). This might be due to the presumed high individuality of sexual preferences. Motivational value plays a key role in many addictive disorders (Robinson & Berridge, 1993). However, there is still little research on VSS reward value computation at an individual level and the role of this process in disorder development. Instead, previous studies used categorical approaches comparing preferred with non-preferred VSS (Brand et al., 2016; Ponseti et al., 2006) or explicit with less explicit VSS (Sescousse et al., 2010; Sescousse, Barbalat et al., 2013; Walter et al., 2008). These approaches indicated that the NAcc, caudate nucleus and OFC were involved in coding the predicted reward outcome of sexual stimuli somehow but it still remained unknown whether these structures encode the whole range of individual preference or if their differential activation represents categories of rewarding and non-rewarding VSS. Additionally, these studies used static picture VSS, which are less ecologically valid compared to films (Solano et al., 2020). Thus, I aimed to examine whether NAcc, caudate nucleus and OFC would encode individual reward value when presented with a range of highly attractive, highly ecologically valid VSS. Since altered striatal reactivity to VSS and cues predicting VSS has been found in connection with compulsive sexual behavior before (Brand et al., 2016; Kühn & Gallinat, 2014; Seok & Sohn, 2015; Voon et al., 2014), I was interested in whether striatal coding of motivational value may be involved here as well. In fact, two studies, that manipulated immediate VSS preference (Brand et al., 2016) and anticipated VSS magnitude (Gola et al., 2017) found striatal value responses positively correlated with PPU symptoms. These are a first indicator, that reward value prediction for VSS may be more refined in persons with more PPU symptoms.

Therefore, individual reward valuation is likely positively associated with PPU severity. Using functional magnetic resonance (fMRI) imaging, I investigated whether neural reactivity to VSS in known reward-related regions (NAcc, caudate nucleus, OFC) is positively associated with individual VSS ratings and if this association is related to self-reported PPU.

While the aim of the first project was to elucidate motivational reward value in the human brain, in my second project I wanted to gain a better understanding of motivational salience in a learning context. Learned motivational salience can be seen as a common factor between reward and fear learning, conceptualized as appetitive and aversive conditioning. Aversive conditioning (or fear conditioning) is conceptualized similarly to appetitive conditioning, with a CS+ and UCS, the main difference being that the UCS is not a reward but an aversive stimulus (e.g. electric shock). The neural correlates of aversive conditioning have been researched extensively in human neuroimaging, so many fMRI results and some meta-analyses on the 'fear network' exist (for reviews see Etkin & Wager, 2007; Fullana et al., 2016; Mechias et al., 2010; Sehlmeyer et al., 2009). Since fMRI studies on appetitive conditioning have begun to accumulate (for reviews see Averbeck & Costa, 2017; Chase et al., 2015; Martin-Soelch et al., 2007), it has become increasingly apparent that the findings from aversive and appetitive conditioning seem similar. Common regions often emerge from separate meta-analyses of responses to a CS+ compared to a CS- in aversive (Etkin & Wager, 2007; Fullana et al., 2016; Mechias et al., 2010) and appetitive (Chase et al., 2015) conditioning, including the NAcc, caudate nucleus, putamen, and amygdala. The cerebellum has been shown associated with aversive conditioning in human data (Ernst et al., 2019; Fullana et al., 2016) appetitive prediction in animal data (Heffley & Hull, 2019), and may thus also be crucial for many different types of outcome prediction (Popa & Ebner, 2018). So, these regions seem to be involved in both aversive and appetitive learning, based on qualitative comparison of empirical data as well as theoretical models. Based on these anatomical overlaps, it is assumed that the concepts 'fear network' and 'reward network' share mesolimbic dopamine pathways and thus may share a common functional basis in a motivational system related to learning motivational salience (Menon & Uddin, 2010; Moscarello & LeDoux, 2013; Seeley et al., 2007; Stefanova et al., 2020). But importantly, this hypothesis is mostly based on qualitative literature reviews. Apart from a very recent meta-analysis on prediction errors (Corlett et al., 2022) few studies have compared the fMRI data from aversive and appetitive paradigms empirically to identify neural commonalities of predicting rewards (appetitive CS+) and

predicting punishment (aversive CS-). In my project, I aimed to investigate how similar a neural activation pattern from an aversive conditioning meta-analysis (aversive CS+ > aversive CS-) is to appetitive conditioning fMRI data (appetitive CS+ > appetitive CS-), to identify potential functional correlates of CS+ motivational salience. I tested pattern expression (i.e. cosine similarity; Bobadilla-Suarez et al., 2020; Weaverdyck et al., 2020) in three independent appetitive conditioning datasets. I expected that the neural response patterns related to aversive and appetitive conditioning would be significantly similar across the different samples. Toward this goal, I tested aversive pattern expression in the appetitive contrasts for the whole brain as well as in common conditioning regions. In this dissertation, I focus on NAcc, caudate nucleus, putamen, amygdala and cerebellum but also tested the thalamus and insula.

3. Summary of Published, Peer-Reviewed Articles

Study I: Subjective reward value of visual sexual stimuli is coded in human striatum and orbitofrontal cortex (Klein et al., 2020)

To answer the first research question, 72 heterosexual and bisexual men took part in an experiment were they were shown visual sexual stimuli (VSS) during an fMRI scan. The VSS shown in the scanner were short film clips (6s) presented without sound, showing at least one woman with a partner during sexual activity. Out of 50 VSS, pre-rated by a different sample, 21 VSS with high valence and sexual arousal ratings were chosen for the scanner experiment. Additionally, control film clips of equal length and similar visual but no sexual content (physiotherapeutic non-sexual massages) were rated on the same scales. Out of 50 control clips, 21 with low sexual arousal and medium valence were chosen for the scanner experiment.

This experiment was a sexual incentive delay (SID) task, an adaption from the already established monetary incentive delay task (Knutson et al., 2001). The SID consisted of 63 trials with three conditions (21 x *VSS*, 21 x *Control*, 21 x *Nothing*), lasting about 20 min. Each trial consisted of an anticipation phase and a delivery phase. During the anticipation phase, one of three geometric shapes served as *CueVSS*, *CueControl* or *CueNothing*. Which shape served as which cue was balanced across participants. Participants were instructed about the

associations between cues and film clips and learned them in a practice task before entering the scanner. Each trial of the SID begins with presentation of one of the three cues (4 s) followed by a fixation cross for a variable (1-3 s) interstimulus interval and then a target stimulus (white square) for at minimum 16 ms and at most 750ms. Participants were instructed to press a reaction button every time this target was presented regardless of the cue presented before. Pressing the reaction button while the target was visible resulted in the win of a film clip if a *CueVSS* or a *CueControl* was shown before. Wins were scheduled in advance to ensure a reinforcement rate of about 71% (15 of 21 *VSS* and *Control* trials each). Target presentation time was adjusted in advance and adapted online if necessary to make it easy (long presentation) or difficult (short presentation) to win according to the pre-planned schedule. After target presentation, a fixation cross was presented for a variable (0–2 s) interstimulus interval, followed by the presentation of either a VSS clip (*CueVSS* & fast response), a control film clip (*CueControl* & fast response) or a black screen (*CueVSS* & slow response, *CueControl* & slow response, *CueNothing*) for 6 s. After a variable (2-6 s) inter-trial interval, the next trial started.

After the experiment, individual ratings of the film clips and questionnaire data were collected outside of the scanner. Participants rated the stimuli set on valence and sexual arousal. All 21 control and 21 VSS film clips were presented, each followed by two 9-point Likert-type scales. Both scales ranged from '1' (indicating 'very unpleasant' or 'not sexually arousing at all') to '9' (indicating 'very pleasant' or 'very sexually arousing'). The resulting VSS ratings were mean-centered and included as parametric modulators in two separate fMRI first level models. These modulators modelled the *DeliveryVSS* events weighted with the ratings of the clips shown at the respective times. Thus, two parametric modulators were examined: '*DeliveryVSS* x valence' and '*DeliveryVSS* x sexual arousal'. A positive result in one of these modulators would indicate a positive correlation between individual VSS ratings and hemodynamic responses during viewing the respective VSS. To assess problematic pornography use (PPU), participants digitally filled out the German versions of the short Internet Addiction Test modified for sexual content, which can result in a total score of 12–60 from 12 items (s-IATsex; Laier et al., 2014). The items measure the experience of negative consequences and a loss of control regarding participants' use of online sexual content as well as craving, social problems, preoccupation and mood regulation. With the fMRI data, I computed two linear regressions with the s-IATsex total scores as predictor and the parametric modulators as outcome at group

level. A positive result here would indicate that the more PPU symptoms a participant reports (s-IATsex total score), the closer the association between individual ratings and hemodynamic responses during VSS viewing.

As expected, s-IATsex scores indicated mainly subclinical PPU symptoms in this sample (mean score = 20.42, SD = 8.03, range = 12-56). VSS valence and sexual arousal ratings were associated with each other ($r = .659$, $p<.001$) but not with the s-IATsex scores. The group-level fMRI analyses revealed that hemodynamic responses during VSS viewing were correlated with both valence and sexual arousal ratings in two regions. The higher the ratings of a VSS clip, the higher hemodynamic responses were in bilateral NAcc and bilateral caudate nucleus during viewing of this VSS clip. Hemodynamic responses in bilateral OFC correlated with valence ratings only, not sexual arousal. The regression results showed that the s-IATsex score was positively associated with the correlation of BOLD response and sexual arousal ratings in left NAcc and bilateral caudate nucleus. There were no significant results with the valence ratings but a trend towards similar positive correlation effects was observed. Thus, the more problems with internet pornography use a subject reported, the stronger the association between hemodynamic responses during VSS viewing and sexual arousal rating of the respective VSS clip in NAcc and caudate nucleus.

The main results extend past studies, where NAcc activity during VSS processing has been associated with subjective aspects of reward (Sabatinelli et al., 2007; Sescousse et al., 2010; Sescousse, Caldú et al., 2013) such as dichotomous VSS intensity or preference categories (Sescousse et al., 2010; Walter et al., 2008) to a linear scaling of NAcc activity along individual VSS reward prediction. This activity may code both motivational value and salience (Cloutier et al., 2008; Cooper & Knutson, 2008; Gerdes et al., 2010; Knutson et al., 2001) of VSS. The caudate nucleus has been implicated in general VSS processing (Graf et al., 2014; Metzger et al., 2010; Seok & Sohn, 2015), but linear scaling effects with individual ratings have not been reported before. The results suggest that caudate nucleus activity aligns with individual liking, which might reflect the predicted reward of a VSS-directed action (e.g. sexual activity) based on the VSS preferences. Third, the OFC has been long-established as an indicator of hedonic valence of many rewards (Peters & Büchel, 2010). This expands the previous VSS literature from the OFC coding categorical intensity (Sescousse et al., 2010; Walter et al., 2008) to a linear relationship between ratings and neural activity among generally pleasant VSS. Since

the striatal regions were related to both sexual arousal and valence ratings while the OFC was exclusively related to valence, this region may reflect the hedonic component of predicted VSS reward more as opposed to motivational components. The strength of association between striatal activity and ratings was greater in participants who reported more PPU symptoms. The individual differences in striatal reward prediction might represent a mechanism that mediates addictive VSS use experienced by some individuals. This extends past studies, where PPU has been linked to a higher striatal response to VSS as compared to a control or non-preferred condition (Brand et al., 2016; Voon et al., 2014). One study, using an SID task with cues containing information about VSS value, found increased NAcc activity associated with increased PPU during the anticipation phase (Gola et al., 2017). This relationship was modulated by dichotomous VSS intensity (explicit vs non-explicit). The results support the notion that motivational value prediction signals in NAcc and caudate differentiate more strongly between differently preferred stimuli, the more PPU symptoms a subject experiences.

Study II: Similarity between neural activity during appetitive conditioning and neural signature of fear conditioning (Klein et al., 2022)

To address the second research question, I re-analyzed three previously published appetitive conditioning datasets in relation to an aversive conditioning meta-analysis. Those three datasets were the *Active Learning/Homogeneous Sample* (n = 29 [men only], Kruse et al., 2017), the *Active Learning/Heterogeneous Sample* (n = 76 [40 women], Kruse et al., 2020) and the *Passive Learning/Heterogeneous Sample* (n = 38 [16 women], Tapia León et al., 2019). Both *Active* samples underwent the same appetitive uninstructed differential conditioning paradigm, adapted from the monetary incentive delay task (Knutson et al., 2001). In each trial, participants were presented with an appetitive CS+ (appCS+) or appetitive CS- (appCS-, blue or yellow rectangle) and then with a target (white square), upon which they were instructed to press a button as quickly as possible. Reactions within target presentation time were rewarded with 50 cents (UCS) only if an appCS+ was presented before the target (timing of the target was predetermined, so that approx. 62% of all appCS+ trials were rewarded). Fast reactions after an appCS- were never rewarded. Participants were instructed to pay attention to any relationships between stimuli before the task and received the money they won after

scanning. The paradigm included 21 appCS+ and 21 appCS- trials. The first two trials were excluded from further analyses, since learning could not have taken place yet, leaving 20 appCS+ and appCS- trials each per subject. In the *Passive Learning/Heterogeneous Sample* (Tapia León et al., 2019), an instructed differential conditioning paradigm without any behavioral reaction component was used. Participants were presented with an appCS+ or appCS- (blue or yellow rectangles) followed by feedback about reward/no reward. Half of the appCS+ trials were rewarded with 50 cents (UCS) while the appCS- was never rewarded. Participants were instructed about the relationships between appCS and UCS before the task and received the money from the experiment after leaving the scanner. The paradigm included 20 appCS+ trials and 20 appCS- trials.

I used a whole brain aversive conditioning pattern which was the result of a meta-analysis of 27 independent fear conditioning data sets (total subjects N = 677, 54% male; Fullana et al., 2016). The aversive conditioning pattern discriminates within aversive conditioning paradigms between CS+ (avCS+) and CS- (avCS-). Most of the studies included used electric shocks as UCS and simple geometric shapes as avCS. The whole brain map of z-values associated with the difference between avCS+ and avCS- was obtained from Neurovault (https://identifiers.org/neurovault.collection:2472). The z-values in this pattern were treated as a pattern of weights in later analysis. I followed the same analysis steps in each sample: (1) First, I computed pattern expression scores (cosine similarity metric) in the whole brain appCS+ > appCS- contrast images. (2) Second, I computed the pattern expression separately for ROIs NAcc, caudate nucleus, putamen, amygdala, insula, thalamus and cerebellum. (3) Finally, I computed pattern expression scores in the separate appCS+ and appCS- activation data, which was then used in a classification analysis to test if I can distinguish appCS+ from appCS- condition based on these scores.

The avCS+ > avCS- pattern was found significantly similar to appCS+ > appCS- contrast images in every sample. The aversive pattern was also more similar to the appetitive conditioning data than other patterns related to cognitive demands or emotional arousal. This similarity was not only present when looking at whole brain activation but also in smaller ROIs. Similarity was highest in NAcc, caudate nucleus and putamen, with more moderate but still significant similarity in cerebellum and amygdala. Furthermore, the aversive pattern could also accurately distinguish appCS+ from appCS- activation in every sample.

This study enabled quantification of the long-assumed similarity of aversive and appetitive learning processes at a neural level (Menon & Uddin, 2010; Moscarello & LeDoux, 2013; Seeley et al., 2007; Stefanova et al., 2020). The results suggest that the differential activation during avCS+ > avCS- contains neural activation independent of UCS valence. This common activation might represent the acquired salience of both avCS+ and appCS+ and thus the motivational salience component of the prediction error (Ogawa et al., 2013; Treviño, 2015). The results mirror meta-analytical regional overlap in activation related to both negative and positive affective processing (Satpute et al., 2015) and appetitive and aversive prediction errors (Corlett et al., 2022). Similarity was especially high in the striatal regions, which fits well with the previously discussed roles of NAcc in reward and loss anticipation (Oldham et al., 2018), caudate nucleus in processing motivational values of actions (Balleine & O'Doherty, 2010) and the putamen in stimulus-response learning (Everitt & Robbins, 2013).

4. Discussion

In this dissertation, I examined the neural correlates of subjective reward values of VSS and the neural commonalities of reward and fear learning. Both research questions concerned the dopaminergic system with a focus on the striatal subregions. Motivational value and salience have a direct influence on reward learning in that they contribute to the size of the reward prediction error, a central factor in learning processes. Our individual reward learning past in turn affects the value and salience that we assign to rewards in the present.

With the first study, I was able to show how VSS reward values are represented in NAcc, caudate nucleus and OFC activity along a linear scale of self-report ratings. These VSS were highly attractive films, kept as close as possible to regular viewing habits of the participants to increase ecological validity. Other VSS studies before ours had only reported value responses to dichotomous categories of preferred and non-preferred static picture VSS (Brand et al., 2016; Ponseti et al., 2006; Sescousse et al., 2010; Sescousse, Barbalat et al., 2013; Walter et al., 2008). In this study, I found that NAcc, caudate nucleus and OFC linearly code the range of value even in this selection of generally highly valued stimuli. The results also corroborated regional specificity for the different components of subjective value, as the OFC was only involved in pleasantness ratings, not the motivational component reflected by the sexual arousal ratings. The fact that the strength of motivational value responses in NAcc and caudate nucleus was also associated with symptoms of PPU indicated increased differentiation of VSS

value in problematic use. This is in line with the incentive sensitization theory of addiction (Robinson & Berridge, 1993), with the addition that sensitization may not apply to addiction-related stimuli all-over but to the value computation of these stimuli. This raises interesting questions concerning disorder development which are further discussed under clinical implications.

While the first study focused on motivational value of rewards, the second explored motivational salience through the commonalities of appetitive and aversive learning processes. The second study showed that appetitive CS+ elicit similar brain activation patterns to aversive CS+. When restricting activation data to NAcc, caudate nucleus or putamen, similarities grew even larger. These regions could thus be especially crucial for learning the motivational salience of stimuli, both appetitive and aversive. With these results, I could show directly in human fMRI data what has been mostly examined in animal data thus far. Direct cell recording studies have shown that appetitive and aversive CS+ may evoke distinct neural responses, but they are often co-localized in the same anatomical areas (O'Neill et al., 2018; Shabel & Janak, 2009; Tye et al., 2010; Xiu et al., 2014). Furthermore, the study showed feasibility of a novel cross-paradigm integration approach used to empirically assess commonalities between different paradigms. Data integration across studies is becoming an increasingly essential analysis tool due to the exponential increase in fMRI publications the difficulties associated with collecting large datasets at single institutions. Using this method, remarkably high neural similarity with the aversive activation pattern in every appetitive sample was found. This enables conclusions about, first, the neural similarity of aversive and appetitive learning itself but second, also about the generalizability of this similarity.

As a final illustration of the scientific value added by the combination of my doctorate projects, the NAcc is a good example since it has already been studied very extensively. The classical view of the NAcc is that the phasic dopamine activity observed in response to reward or reward-predictive stimuli represents a (learned) motivational value, which is sensitive to changes to the outcome, i.e. a signal to approach these stimuli (Flagel et al., 2010). For aversive conditioning, NAcc dopamine seems to convey a motivational signal to avoid the aversive UCS if that is possible (Gentry et al., 2019). One explanation for this is that different neuronal populations in the ventral tegmental areal affect different anatomical subdivisions of the NAcc, one conveying the unsigned motivational salience signal (NAcc core) and one the

signed motivational value signal (NAcc shell; Bromberg-Martin et al., 2010). Animal studies suggest that the medial NAcc shell functions as an 'affective keyboard', generating positively valenced emotions at its anterior end and increasingly negatively valenced emotions towards the posterior end (Berridge & Kringelbach, 2015). My results support the existence of both motivational value and motivational salience signals in this region. NAcc activation in association with VSS value as well as in response to the CS+ during appetitive conditioning indicates the motivational value of VSS and CS+ (likely originating from the NAcc shell). The commonality found between appetitive and aversive CS+ may in turn convey the motivational salience of both CS+ (likely originating from the NAcc core). Thus, the combination of my two projects corroborates the dual-function view of this region which had mostly been based on animal literature thus far.

Clinical Implications & Future Directions

The results of both studies may have important clinical implications. Concerning the increased differentiation of motivational value signals with increasing addictive symptoms suggested by study 1, this might be connected to an increase in time spent searching for highly valued VSS, which could lead to issues in personal or professional life because of this behavior. Regarding the development timeline of the problematic behavior, it would be interesting if the differences between values got larger over time, requiring the search for and use of more and more highly preferred VSS to obtain the same reward. To further elucidate these mechanisms and find out whether they are precursor or result of the addictive behavior, longitudinal studies are needed. In the second study, I did not examine any measures of risk behaviors but altered aversive and appetitive conditioning are considered the basis for psychological disorders characterized by excessive avoidance and approach behavior, respectively (Duits et al., 2015; Martin-Soelch et al., 2007), for example addictions and anxiety disorders. Still very little is known about commonalities and overlaps between these disorder categories. This study provides proof of concept for an approach which facilitates finding commonalities in such separate concepts. Further integration of data across more different affective learning paradigms and Research Domain Criteria domains – and across patient samples - may help fill these knowledge gaps and further pave the way towards transdiagnostic biomarkers (Insel, 2014; Woo et al., 2017).

Summary

This dissertation could advance the knowledge concerning motivational processing in the brain, particularly concerning subjective value of highly individualized rewards and the commonalities of appetitive and aversive predictive processing. Altogether, 10 articles concerning reward processing and learning were published in peer-reviewed journals, 3 of those as first author, 1 in shared first-authorship and 6 as co-author. This work shows that the neural responses in NAcc, caudate, and OFC closely code individual preferences among highly rewarding sexual stimuli, with the striatal regions being especially important for motivational value. The association of PPU with the correspondence between individual ratings and neural activity might allow new insights in the neurobiological mechanisms underlying the development of compulsive sexual behavior disorder. The work further demonstrates the similarity of aversive and appetitive prediction at fMRI pattern level – especially in dopaminergic striatal circuits - across multiple independent appetitive datasets. This quantifies what previously where qualitative assumptions concerning systems of motivational salience and opens up interesting implications for etiological models of fear- and reward-related disorders. Finally, the novel approach of quantitative cross-paradigm integration presents an opportunity to integrate rather than compare past findings with current studies and thus make better use of the always growing body of fMRI studies.

5. References

Arias-Carrión, O., Stamelou, M., Murillo-Rodríguez, E., Menéndez-González, M., & Pöppel, E. (2010). Dopaminergic reward system: A short integrative review. *International Archives of Medicine*, *3*, 24. https://doi.org/10.1186/1755-7682-3-24

Averbeck, B. B., & Costa, V. D. (2017). Motivational neural circuits underlying reinforcement learning. *Nature Neuroscience*, *20*(4), 505–512. https://doi.org/10.1038/nn.4506

Baeyens, F., Eelen, P., & van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: a case for unaware affective-evaluative learning. Advance online publication. https://doi.org/10.1080/02699939008406760 (Cognition & Emotion, 4(1), 3-18).

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. https://doi.org/10.1038/npp.2009.131

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron*, *86*(3), 646–664. https://doi.org/10.1016/j.neuron.2015.02.018

Bobadilla-Suarez, S., Ahlheim, C., Mehrotra, A., Panos, A., & Love, B. C. (2020). Measures of Neural Similarity. *Computational Brain & Behavior*, *3*(4), 369–383. https://doi.org/10.1007/s42113-019-00068-5

Brand, M., Snagowski, J., Laier, C., & Maderwald, S. (2016). Ventral striatum activity when watching preferred pornographic pictures is correlated with symptoms of Internet pornography addiction. *NeuroImage*, *129*, 224–232. https://doi.org/10.1016/j.neuroimage.2016.01.033

Brand, M., Wegmann, E., Stark, R., Müller, A., Wölfling, K., Robbins, T. W., & Potenza, M. N. (2019). The interaction of Person-Affect-Cognition-Execution (I-PACE) model for addictive behaviors: update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors.

Neuroscience and Biobehavioral Reviews, *104*, 1–10. https://doi.org/10.1016/j.neubiorev.2019.06.032

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron*, *68*(5), 815–834. https://doi.org/10.1016/j.neuron.2010.11.022

Carta, I., Chen, C. H., Schott, A. L., Dorizan, S., & Khodakhah, K. (2019). Cerebellar modulation of the reward circuitry and social behavior. *Science (New York, N.Y.)*, *363*(6424). https://doi.org/10.1126/science.aav0581

Chase, H. W., Kumar, P., Eickhoff, S., & Dombrovski, A. Y. (2015). Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*, *15*(2), 435–459. https://doi.org/10.3758/s13415-015-0338-7

Ciocchi, S., Herry, C., Grenier, F., Wolff, S. B. E., Letzkus, J. J., Vlachos, I., Ehrlich, I., Sprengel, R., Deisseroth, K., Stadler, M. B., Müller, C., & Lüthi, A. (2010). Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature*, *468*(7321), 277–282. https://doi.org/10.1038/nature09559

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, *9*(9), 1289–1302. https://doi.org/10.1093/scan/nst106

Cloutier, J., Heatherton, T. F., Whalen, P. J., & Kelley, W. M. (2008). Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. *Journal of Cognitive Neuroscience*, *20*(6), 941–951. https://doi.org/10.1162/jocn.2008.20062

Cooper, J. C., & Knutson, B. (2008). Valence and salience contribute to nucleus accumbens activation. *NeuroImage*, *39*(1), 538–547. https://doi.org/10.1016/j.neuroimage.2007.08.009

Corlett, P. R., Mollick, J. A., & Kober, H. (2022). Meta-analysis of human prediction error for incentives, perception, cognition, and action. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*. Advance online publication. https://doi.org/10.1038/s41386-021-01264-3

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Annals of the New York Academy of Sciences*, *1104*, 70–88. https://doi.org/10.1196/annals.1390.002

Duits, P., Cath, D. C., Lissek, S., Hox, J. J., Hamm, A. O., Engelhard, I. M., van den Hout, M. A., & Baas, J. M. P. (2015). Updated meta-analysis of classical fear conditioning in the anxiety disorders. Advance online publication. https://doi.org/10.1002/DA.22353 (Depression and Anxiety, 32(4), 239-253).

Ernst, T. M., Brol, A. E., Gratz, M., Ritter, C., Bingel, U., Schlamann, M., Maderwald, S., Quick, H. H., Merz, C. J., & Timmann, D. (2019). The cerebellum is involved in processing of predictions and prediction errors in a fear conditioning paradigm. *ELife*, *8.* https://doi.org/10.7554/eLife.46831

Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *The American Journal of Psychiatry*, *164*(10), 1476–1488. https://doi.org/10.1176/appi.ajp.2007.07030504

Everitt, B. J., & Robbins, T. W. (2013). From the ventral to the dorsal striatum: Devolving views of their roles in drug addiction. *Neuroscience and Biobehavioral Reviews*, *37*(9 Pt A), 1946–1954. https://doi.org/10.1016/j.neubiorev.2013.02.010

Flagel, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czuj, A., Willuhn, I., Akers, C. A., Clinton, S. M., Phillips, P. E. M., & Akil, H. (2010). A selective role for dopamine in reward learning. *Nature*, *469*(7328), 53–57. https://doi.org/10.1038/nature09588

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, *21*(4), 500–508. https://doi.org/10.1038/mp.2015.88

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *37*(7), 1297–1310. https://doi.org/10.1016/j.neubiorev.2013.03.023

Gentry, R. N., Schuweiler, D. R., & Roesch, M. R. (2019). Dopamine signals related to appetitive and aversive events in paradigms that manipulate reward and avoidability. *Brain Research*, *1713*, 80–90. https://doi.org/10.1016/j.brainres.2018.10.008

Gerdes, A. B. M., Wieser, M. J., Mühlberger, A., Weyers, P., Alpers, G. W., Plichta, M. M., Breuer, F., & Pauli, P. (2010). Brain activations to emotional pictures are differentially associated with valence and arousal ratings. *Frontiers in Human Neuroscience*, *4*, 175. https://doi.org/10.3389/fnhum.2010.00175

Gola, M., Wordecha, M., Sescousse, G., Lew-Starowicz, M., Kossowski, B., Wypych, M., Makeig, S., Potenza, M. N., & Marchewka, A. (2017). Can pornography be addictive? An fMRI study of men seeking treatment for problematic pornography use. *Neuropsychopharmacology*, *42*(10), 2021–2031. https://doi.org/10.1038/npp.2017.78

Graf, H., Wiegers, M., Metzger, C. D., Walter, M., Grön, G., & Abler, B. (2014). Erotic stimulus processing under amisulpride and reboxetine: A placebo-controlled fMRI study in healthy subjects. *The International Journal of Neuropsychopharmacology*, *18*(2). https://doi.org/10.1093/ijnp/pyu004

Haber, S. N., & Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology*, *35*(1), 4–26. https://doi.org/10.1038/npp.2009.129

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. https://doi.org/10.1126/science.1168450

Heffley, W., & Hull, C. (2019). Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. *ELife*, *8.* https://doi.org/10.7554/eLife.46764

Howard, J. D., & Kahnt, T. (2021). To be specific: The role of orbitofrontal cortex in signaling reward identity. *Behavioral Neuroscience*, *135*(2), 210–217. https://doi.org/10.1037/bne0000455

Insel, T. R. (2014). The NIMH Research Domain Criteria (RDoC) Project: Precision medicine for psychiatry. *The American Journal of Psychiatry*, *171*(4), 395–397. https://doi.org/10.1176/appi.ajp.2014.14020138

Janak, P. H., & Tye, K. M. (2015). From circuits to behaviour in the amygdala. *Nature*, *517*(7534), 284–292. https://doi.org/10.1038/nature14188

Kim, H., Shimojo, S., & O'Doherty, J. P. (2011). Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cerebral Cortex*, *21*(4), 769–776. https://doi.org/10.1093/cercor/bhq145

Kita, H., & Kitai, S. T. (1990). Amygdaloid projections to the frontal cortex and the striatum in the rat. *The Journal of Comparative Neurology*, *298*(1), 40–49. https://doi.org/10.1002/cne.902980104

Klein, S., Krikova, K., Antons, S., Brand, M., Klucken, T., & Stark, R. (in press). Reward Responsiveness, Learning and Valuation Implicated in Problematic Pornography Use – a Research Domain Criteria Perspective. *Current Addiction Reports*.

Klein, S., Kruse, O., Markert, C., Tapia León, I., Strahler, J., & Stark, R. (2020). Subjective reward value of visual sexual stimuli is coded in human striatum and orbitofrontal cortex. *Behavioural Brain Research*, *393*, 112792. https://doi.org/10.1016/j.bbr.2020.112792

Klein, S., Kruse, O., Tapia León, I., Stalder, T., Stark, R., & Klucken, T. (2019). Increased neural reactivity to emotional pictures in men with high hair testosterone concentrations. *Social Cognitive and Affective Neuroscience*, *14*(9), 1009–1016. https://doi.org/10.1093/scan/nsz067

Klein, S., Kruse, O., Tapia León, I., van Oudenhove, L., van 't Hof, S. R., Klucken, T., Wager, T. D., & Stark, R. (2022). Cross-paradigm integration shows a common neural basis for aversive and appetitive conditioning. *NeuroImage*, *263*, 119594. https://doi.org/10.1016/j.neuroimage.2022.119594

Klein-Flügge, M. C., Barron, H. C., Brodersen, K. H., Dolan, R. J., & Behrens, T. E. J. (2013). Segregated encoding of reward-identity and stimulus-reward associations in human orbitofrontal cortex. *The Journal of Neuroscience : The Official Journal of the Society*

for Neuroscience, *33*(7), 3202–3211. https://doi.org/10.1523/JNEUROSCI.2532-12.2013

Klucken, T., Kruse, O., Klein, S., Kampa, M., Tapia León, I., & Stark, R. (2019). The relationship between neuroticism and appetitive conditioning. *Neurobiology of Learning and Memory*, *164*, 107068. https://doi.org/10.1016/j.nlm.2019.107068

Klucken, T., Wehrum, S., Schweckendiek, J., Merz, C. J., Hennig, J., Vaitl, D., & Stark, R. (2013). The 5-HTTLPR polymorphism is associated with altered hemodynamic responses during appetitive conditioning. *Human Brain Mapping*, *34*(10), 2549–2560. https://doi.org/10.1002/hbm.22085

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience*, *21*(16), RC159-RC159. https://doi.org/10.1523/JNEUROSCI.21-16-j0002.2001

Kruse, O., Klein, S., Tapia León, I., Stark, R., & Klucken, T. (2020). Amygdala and nucleus accumbens involvement in appetitive extinction. *Human Brain Mapping*, *41*(7), 1833–1841. https://doi.org/10.1002/hbm.24915

Kruse, O., Tapia León, I., Stark, R., & Klucken, T. (2017). Neural correlates of appetitive extinction in humans. *Social Cognitive and Affective Neuroscience*, *12*(1), 106–115. https://doi.org/10.1093/scan/nsw157

Kühn, S., & Gallinat, J. (2012). The neural correlates of subjective pleasantness. *NeuroImage*, *61*(1), 289–294. https://doi.org/10.1016/j.neuroimage.2012.02.065

Kühn, S., & Gallinat, J. (2014). Brain structure and functional connectivity associated with pornography consumption: The brain on porn. *JAMA Psychiatry*, *71*(7), 827–834. https://doi.org/10.1001/jamapsychiatry.2014.93

LaBar, K. S., Gatenby, J.C., Gore, J. C., LeDoux, J. E., & Phelps, E. A. (1998). Human Amygdala Activation during Conditioned Fear Acquisition and Extinction: a Mixed-Trial fMRI Study. *Neuron*, *20*(5), 937–945. https://doi.org/10.1016/S0896-6273(00)80475-4

Laier, C., Pekal, J., & Brand, M. (2014). Cybersex addiction in heterosexual female users of internet pornography can be explained by gratification hypothesis. *Cyberpsychology,*

Behavior and Social Networking, *17*(8), 505–511. https://doi.org/10.1089/cyber.2013.0396

Lam, J. M., Wächter, T., Globas, C., Karnath, H.-O., & Luft, A. R. (2013). Predictive value and reward in implicit classification learning. *Human Brain Mapping*, *34*(1), 176–185. https://doi.org/10.1002/hbm.21431

Lange, I., Kasanova, Z., Goossens, L., Leibold, N., Zeeuw, C. I. de, van Amelsvoort, T., & Schruers, K. (2015). The anatomy of fear learning in the cerebellum: A systematic meta-analysis. *Neuroscience and Biobehavioral Reviews*, *59*, 83–91. https://doi.org/10.1016/j.neubiorev.2015.09.019

Le, T. M., Zhang, S., Zhornitsky, S., Wang, W., & Li, C.-S. R. (2020). Neural correlates of reward-directed action and inhibition of action. *Cortex*, *123*, 42–56. https://doi.org/10.1016/j.cortex.2019.10.007

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: Evidence from functional neuroimaging. *Neuron*, *64*(3), 431–439. https://doi.org/10.1016/j.neuron.2009.09.040

LeDoux, J. E., Iwata, J., Cicchetti, P., & Reis, D. J. (1988). Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *The Journal of Neuroscience*, *8*(7), 2517–2529. https://doi.org/10.1523/JNEUROSCI.08-07-02517.1988

Luo, S., Ainslie, G., Giragosian, L., & Monterosso, J. R. (2009). Behavioral and neural evidence of incentive bias for immediate rewards relative to preference-matched delayed rewards. *The Journal of Neuroscience*, *29*(47), 14820–14827. https://doi.org/10.1523/JNEUROSCI.4261-09.2009

Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. Advance online publication. https://doi.org/10.1037/H0076778 (Psychological Review, 82(4), 276-298).

Marchner, JR, & Preuschhof, C. (2018). Reward history but not search history explains value-driven attentional capture. *Attention, Perception & Psychophysics*, *80*(6). https://doi.org/10.3758/s13414-018-1513-z

Markert, C., Klein, S., Strahler, J., Kruse, O., & Stark, R. (2021). Sexual incentive delay in the scanner: Sexual cue and reward processing, and links to problematic porn consumption and sexual motivation. *Journal of Behavioral Addictions*, *10*(1), 65–76. https://doi.org/10.1556/2006.2021.00018

Martin-Soelch, C., Linthicum, J., & Ernst, M. (2007). Appetitive conditioning: neural bases and implications for psychopathology. *Neuroscience and Biobehavioral Reviews*, *31*(3). https://doi.org/10.1016/j.neubiorev.2006.11.002

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837–841. https://doi.org/10.1038/nature08028

Mechias, M.-L., Etkin, A., & Kalisch, R. (2010). A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. *NeuroImage*, *49*(2), 1760–1768. https://doi.org/10.1016/j.neuroimage.2009.09.040

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure & Function*, *214*(5-6), 655–667. https://doi.org/10.1007/s00429-010-0262-0

Metereau, E., & Dreher, J.-C. (2015). The medial orbitofrontal cortex encodes a general unsigned value signal during anticipation of both appetitive and aversive events. *Cortex*, *63*, 42–54. https://doi.org/10.1016/j.cortex.2014.08.012

Metzger, C. D., Eckert, U., Steiner, J., Sartorius, A., Buchmann, J. E., Stadler, J., Tempelmann, C., Speck, O., Bogerts, B., Abler, B., & Walter, M. (2010). High field fMRI reveals thalamocortical integration of segregated cognitive and emotional processing in mediodorsal and intralaminar thalamic nuclei. *Frontiers in Neuroanatomy*, *4*, 138. https://doi.org/10.3389/fnana.2010.00138

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience*, *16*(5), 1936–1947. https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996

Moscarello, J. M., & LeDoux, J. E. (2013). The contribution of the amygdala to aversive and appetitive pavlovian processes. *Emotion Review*, *5*(3), 248–253. https://doi.org/10.1177/1754073913477508

Nassar, M. R., McGuire, J. T., Ritz, H., & Kable, J. W. (2019). Dissociable Forms of Uncertainty-Driven Representational Change Across the Human Brain. *The Journal of Neuroscience*, *39*(9), 1688. https://doi.org/10.1523/JNEUROSCI.1713-18.2018

Ogawa, M., van der Meer, M. A. A., Esber, G. R., Cerri, D. H., Stalnaker, T. A., & Schoenbaum, G. (2013). Risk-responsive orbitofrontal neurons track acquired salience. *Neuron*, *77*(2), 251–258. https://doi.org/10.1016/j.neuron.2012.11.006

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., & Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. *Human Brain Mapping*, *39*(8), 3398–3418. https://doi.org/10.1002/hbm.24184

O'Neill, P.-K., Gore, F., & Salzman, C. D. (2018). Basolateral amygdala circuitry in positive and negative valence. *Current Opinion in Neurobiology*, *49*, 175–183. https://doi.org/10.1016/j.conb.2018.02.012

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex.* Oxford University Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Peters, J., & Büchel, C. (2010). Neural representations of subjective reward value. *Behavioural Brain Research*, *213*(2), 135–141. https://doi.org/10.1016/j.bbr.2010.04.031

Pine, A., Seymour, B., Roiser, J. P., Bossaerts, P., Friston, K. J., Curran, H. V., & Dolan, R. J. (2009). Encoding of marginal utility across time in the human brain. *The Journal of Neuroscience*, *29*(30), 9575–9581. https://doi.org/10.1523/JNEUROSCI.1126-09.2009

Plassmann, H., O'Doherty, J. P., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National*

Academy of Sciences of the United States of America, *105*(3), 1050–1054. https://doi.org/10.1073/pnas.0706929105

Ponseti, J., Bosinski, H. A., Wolff, S., Peller, M., Jansen, O., Mehdorn, H. M., Büchel, C., & Siebner, H. R. (2006). A functional endophenotype for sexual orientation in humans. *NeuroImage*, *33*(3), 825–833. https://doi.org/10.1016/j.neuroimage.2006.08.002

Popa, L. S., & Ebner, T. J. (2018). Cerebellum, Predictions and Errors. *Frontiers in Cellular Neuroscience*, *12*, 524. https://doi.org/10.3389/fncel.2018.00524

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research. Brain Research Reviews*, *18*(3), 247–291. https://doi.org/10.1016/0165-0173(93)90013-P

Rolls, E. T., Grabenhorst, F., & Parris, B. A. (2008). Warm pleasant feelings in the brain. *NeuroImage*, *41*(4), 1504–1513. https://doi.org/10.1016/j.neuroimage.2008.03.005

Rolls, E. T., & McCabe, C. (2007). Enhanced affective brain representations of chocolate in cravers vs. Non-cravers. *The European Journal of Neuroscience*, *26*(4), 1067–1076. https://doi.org/10.1111/j.1460-9568.2007.05724.x

Royet, J.-P., Plailly, J., Delon-Martin, C., Kareken, D. A., & Segebarth, C. (2003). Fmri of emotional responses to odors: Influence of hedonic valence and judgment, handedness, and gender. *NeuroImage*, *20*(2), 713–728. https://doi.org/10.1016/S1053-8119(03)00388-4

Sabatinelli, D., Bradley, M. M., Lang, P. J., Costa, V. D., & Versace, F. (2007). Pleasure rather than salience activates human nucleus accumbens and medial prefrontal cortex. *Journal of Neurophysiology*, *98*(3), 1374–1379. https://doi.org/10.1152/jn.00230.2007

Satpute, A. B., Kang, J., Bickart, K. C., Yardley, H., Wager, T. D., & Barrett, L. F. (2015). Involvement of Sensory Regions in Affective Experience: A Meta-Analysis. *Frontiers in Psychology*, *6*, 1860. https://doi.org/10.3389/fpsyg.2015.01860

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.)*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. *Physiological Reviews*, *95*(3), 853–951. https://doi.org/10.1152/physrev.00023.2014

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience*, *27*(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Sehlmeyer, C., Schöning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., & Konrad, C. (2009). Human fear conditioning and extinction in neuroimaging: A systematic review. *PloS One*, *4*(6), e5865. https://doi.org/10.1371/journal.pone.0005865

Seok, J.-W., & Sohn, J.-H. (2015). Neural substrates of sexual desire in individuals with problematic hypersexual behavior. *Frontiers in Behavioral Neuroscience*, *9*, 321. https://doi.org/10.3389/fnbeh.2015.00321

Sescousse, G., Barbalat, G., Domenech, P., & Dreher, J.-C. (2013). Imbalance in the sensitivity to different types of rewards in pathological gambling. *Brain*, *136*(Pt 8), 2527–2538. https://doi.org/10.1093/brain/awt126

Sescousse, G., Caldú, X., Segura, B., & Dreher, J.-C. (2013). Processing of primary and secondary rewards: A quantitative meta-analysis and review of human functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *37*(4), 681–696. https://doi.org/10.1016/j.neubiorev.2013.02.002

Sescousse, G., Redouté, J., & Dreher, J.-C. (2010). The architecture of reward value coding in the human orbitofrontal cortex. *The Journal of Neuroscience*, *30*(39), 13095–13104. https://doi.org/10.1523/JNEUROSCI.3501-10.2010

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., & Frackowiak, R. S. (2004). Temporal difference models describe higher-order

learning in humans. *Nature*, *429*(6992), 664–667. https://doi.org/10.1038/nature02581

Shabel, S. J., & Janak, P. H. (2009). Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(35), 15031–15036. https://doi.org/10.1073/pnas.0905580106

Skinner, B. F. (1937). Two Types of Conditioned Reflex: A Reply to Konorski and Miller. *The Journal of General Psychology*, *16*(1), 272–279. https://doi.org/10.1080/00221309.1937.9917951

Solano, I., Eaton, N. R., & O'Leary, K. D. (2020). Pornography consumption, modality and function in a large internet sample. *Journal of Sex Research*, *57*(1), 92–103. https://doi.org/10.1080/00224499.2018.1532488

Stark, R., Klein, S., Kruse, O., Walter, B., Strahler, J., & Markert, C. (2022). Acute stress increases incentive of sexual cues. *Journal of Behavioral Addictions*.

Stark, R., Klein, S., Kruse, O., Weygandt, M., Leufgens, L. K., Schweckendiek, J., & Strahler, J. (2019). No sex difference found: Cues of sexual stimuli activate the reward system in both sexes. *Neuroscience*, *416*, 63–73. https://doi.org/10.1016/j.neuroscience.2019.07.049

Stefanova, E., Dubljević, O., Herbert, C., Fairfield, B., Schroeter, M. L., Stern, E. R., Urben, S., Derntl, B., Wiebking, C., Brown, C., Drach-Zahavy, A., Kathrin Loeffler, L. A., Albrecht, F., Palumbo, R., Boutros, S. W., Raber, J., & Lowe, L. (2020). Anticipatory feelings: Neural correlates and linguistic markers. *Neuroscience and Biobehavioral Reviews*, *113*, 308–324. https://doi.org/10.1016/j.neubiorev.2020.02.015

Sutton, R. S., & Barto, A. G. (1987). *A temporal-difference model of classical conditioning.*

Tapia León, I., Kruse, O., Stark, R., & Klucken, T. (2019). Relationship of sensation seeking with the neural correlates of appetitive conditioning. *Social Cognitive and Affective Neuroscience*, *14*(7), 769–775. https://doi.org/10.1093/scan/nsz046

Tovote, P., Fadok, J. P., & Lüthi, A. (2015). Neuronal circuits for fear and anxiety. *Nature Reviews. Neuroscience*, *16*(6), 317–331. https://doi.org/10.1038/nrn3945

Treviño, M. (2015). Associative Learning Through Acquired Salience. *Frontiers in Behavioral Neuroscience*, *9*, 353. https://doi.org/10.3389/fnbeh.2015.00353

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *The European Journal of Neuroscience*, *29*(11), 2225–2232. https://doi.org/10.1111/j.1460-9568.2009.06796.x

Tye, K. M., Cone, J. J., Schairer, W. W., & Janak, P. H. (2010). Amygdala Neural Encoding of the Absence of Reward during Extinction. *The Journal of Neuroscience*, *30*(1), 116–125. https://doi.org/10.1523/JNEUROSCI.4240-09.2010

van 't Hof, S. R., van Oudenhove, L., Janssen, E., Klein, S., Reddan, M. C., Kragel, P. A., Stark, R., & Wager, T. D. (2021). The Brain Activation-Based Sexual Image Classifier (BASIC): A Sensitive and Specific fMRI Activity Pattern for Sexual Image Processing. *Cerebral Cortex.* Advance online publication. https://doi.org/10.1093/cercor/bhab397

Voon, V., Mole, T. B., Banca, P., Porter, L., Morris, L. S., Mitchell, S., Lapa, T. R., Karr, J., Harrison, N. A., Potenza, M. N., & Irvine, M. (2014). Neural correlates of sexual cue reactivity in individuals with and without compulsive sexual behaviours. *PloS One*, *9*(7), e102419. https://doi.org/10.1371/journal.pone.0102419

Walter, M., Bermpohl, F., Mouras, H., Schiltz, K., Tempelmann, C., Rotte, M., Heinze, H. J., Bogerts, B., & Northoff, G. (2008). Distinguishing specific sexual and general emotional effects in fMRI-subcortical and cortical arousal during erotic picture viewing. *NeuroImage*, *40*(4), 1482–1494. https://doi.org/10.1016/j.neuroimage.2008.01.040

Warlow, S. M., Robinson, M. J.F., & Berridge, K. C. (2017). Optogenetic Central Amygdala Stimulation Intensifies and Narrows Motivation for Cocaine. *The Journal of Neuroscience*, *37*(35), 8330–8348. https://doi.org/10.1523/JNEUROSCI.3141-16.2017

Weaverdyck, M. E., Lieberman, M. D., & Parkinson, C. (2020). Tools of the Trade. Multivoxel pattern analysis in fMRI: A practical introduction for social and affective neuroscientists. *Social Cognitive and Affective Neuroscience*, *15*(4), 487–509. https://doi.org/10.1093/scan/nsaa057

Wilson, R. P., Colizzi, M., Bossong, M. G., Allen, P., Kempton, M., & Bhattacharyya, S. (2018). The neural substrate of reward anticipation in health: A meta-analysis of fMRI findings

in the monetary incentive delay task. *Neuropsychology Review*, *28*(4), 496–506. https://doi.org/10.1007/s11065-018-9385-5

Wimmer, E. G., & Büchel, C. (2019). Learning of distant state predictions by the orbitofrontal cortex in humans. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-10597-z

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience*, *20*(3), 365–377. https://doi.org/10.1038/nn.4478

Xiu, J., Zhang, Q., Zhou, T., Zhou, T.-t., Chen, Y., & Hu, H. (2014). Visualizing an emotional valence map in the limbic forebrain by TAI-FISH. *Nature Neuroscience*, *17*(11), 1552–1559. https://doi.org/10.1038/nn.3813

Zhang, J., Berridge, K. C., Tindell, A. J., Smith, K. S., & Aldridge, J. W. (2009). A neural computational model of incentive salience. *PLoS Computational Biology*, *5*(7). https://doi.org/10.1371/journal.pcbi.1000437

II Publications

Publication 1: Subjective reward value of visual sexual stimuli is coded in human striatum and orbitofrontal cortex

Behavioural Brain Research 393 (2020) 112792

Contents lists available at ScienceDirect

Behavioural Brain Research

journal homepage: www.elsevier.com/locate/bbr

Research report

Subjective reward value of visual sexual stimuli is coded in human striatum and orbitofrontal cortex

Sanja Klein[a,b,d,*], Onno Kruse[a,b,c], Charlotte Markert[a,b,d], Isabell Tapia León[a,b,c], Jana Strahler[a,b], Rudolf Stark[a,b,d]

[a] Department of Psychotherapy and Systems Neuroscience, Justus Liebig University, 35394 Giessen, Germany
[b] Bender Institute for Neuroimaging (BION), Justus Liebig University, 35394 Giessen, Germany
[c] Clinical Psychology, University Siegen, 57076 Siegen, Germany
[d] Center of Mind, Brain and Behavior, Universities of Marburg and Giessen, 35032 Marburg, Germany

ARTICLE INFO

Keywords:
fMRI
OFC
Pornography
Reward value
Striatum

ABSTRACT

Human neuroimaging research suggests the existence of one core network for the subjective valuation of rewards, including the striatum and orbitofrontal cortex. However, there is little research on the neural representation of subjective reward values of visual sexual stimuli (VSS) and on the role of these subjective valuations in the development of related addictive behaviors. Here, we investigate how neural reactivity to VSS is connected to individual preference using functional magnetic resonance imaging (fMRI). During the fMRI scan, 72 men viewed different VSS film clips. Ratings regarding valence and sexual arousal were collected and used as parametric modulators in the fMRI analysis. Subjects also filled out questionnaires on self-reported symptoms of problematic pornography use (PPU). Firstly, we found that neural reactivity towards VSS clips in the nucleus accumbens, caudate nucleus and orbitofrontal cortex was positively correlated with individual ratings of the respective VSS in all subjects. Second, the strength of the association between neural activity and sexual arousal ratings was positively correlated with self-reported symptoms of PPU. The first result suggests a precise appraisal of VSS according to individual preferences in established reward valuation regions. Secondly, stronger neural differentiation based on preference in participants with more PPU symptoms indicates an increased importance of VSS/preference fit in these individuals. This heightened correspondence between individual liking and neural activity may facilitate PPU development by increased signaling of incentive salience, thus boosting motivation to seek out and respond to these preferred stimuli.

1. Introduction

A reward's subjective value affects its potential to induce approach, reinforce behavior, and elicit positive emotions. The subjective value is influenced by physical reward properties like magnitude, probability or delay, as well as by subjective preferences, which can be measured by pleasantness or arousal ratings [1]. The subjective value thus includes a hedonic and motivational component, widely termed incentive salience [2]. In contrast to many other rewarding stimuli, research on the neural representation of subjective values for visual sexual stimuli (VSS) has remained largely unexplored. This is surprising since VSS have been widely established as an effective type of reward and shown to recruit neurocircuitries similar to other rewarding stimuli [3,4]. Since problems related to excessive VSS use are wide-spread [5] and have recently been clinically formalized in the International Classification of Disease (ICD-11; [6]), it is all the more relevant to investigate the role of subjective VSS value coding in this context.

Striatal regions have been identified as part of a general reward valuation network in the human brain [7–9], specifically, the nucleus accumbens (NAcc), as part of the ventral striatum, and the caudate nucleus as part of the dorsal (dorsomedial) striatum [10]. Neural activity in these structures has been associated with similar but also distinct functional processes. The NAcc responds to rewarding stimuli as well as stimuli that signal a reward, i.e. stimuli which have incentive salience [11,12]. NAcc activity scales up with increasing magnitude of monetary reward [13] and corresponds to individual ratings of attractive faces [14,15] and tactile warmth [16]. Thus, the NAcc is implicated in coding the subjective value of a rewarding stimulus. The caudate nucleus is involved in preparing and taking actions to obtain a reward [17,18]. Caudate activity is increased when participants actively choose

* Corresponding author at: Department of Psychotherapy and Systems Neuroscience, Justus Liebig University, 35394 Giessen, Germany.
E-mail address: Sanja.Klein@psychol.uni-giessen.de (S. Klein).

https://doi.org/10.1016/j.bbr.2020.112792
Received 6 March 2020; Received in revised form 23 June 2020; Accepted 24 June 2020
Available online 27 June 2020

a monetary reward that is higher in magnitude or less delayed in its delivery [19–21]. This indicates that the caudate nucleus encodes the value of a goal-directed action based on the expected outcome. Activity in the putamen, also being part of the dorsal (dorsolateral) striatum appears to reflect a different process. This structure is involved in stimulus-response-learning and habit development, which is no longer influenced by the value of the outcome [17]. In addition to these striatal regions, the orbitofrontal cortex (OFC) is also implicated in value-related processing [8,9]. Previous research has shown activity in the OFC to scale with individual pleasantness reports for food [22–24], odors [25], attractive faces [14,15] and with the expected magnitude of money and juice rewards [26]. Accordingly, the OFC is thought to encode the hedonic component of subjective reward value, but not necessarily the motivational component contained in the striatal response.

Meta-analyses have identified NAcc, caudate nucleus and OFC as more active when participants view VSS compared to control stimuli, indicating similar general processing of VSS and other rewards [3,4,27]. To better understand subjective VSS value coding however, comparisons between VSS with varying values are necessary. Studies using categorical approaches have found increasing evidence for subjective VSS value coding in these regions. Neural activity in the NAcc is increased in response to preferred VSS as compared to non-preferred VSS [28]. In some studies this activity is also associated with sexual arousal [29] or pleasantness ratings [30,31]. Caudate activity corresponds to sexual arousal [32] and OFC activity to the pleasantness ratings [30] of different VSS categories. Importantly however, all of these studies used comparisons between preferred and non-preferred VSS [28,29] or more and less explicit VSS [30–32], based on pre-defined general appraisals. With this approach, it remains unknown, if the neural structures encode a finely graduated range of individual preferences or if the activity difference only represents dichotomous categories of generally rewarding and non-rewarding VSS. Using individual continuous ratings with generally preferred VSS to reflect the variance of subjective values would be advantageous. This individual correlational approach would represent the personal experiences more accurately than comparing extreme ratings induced by pre-defined categories and enable researchers to examine previously uninvestigated individual differences in pleasantness and sexual arousal regarding the neural representation of VSS reward value. A further shortcoming of previous studies was the use of static picture stimuli. A recent representative study has shown that videos are the predominant pornography used in everyday life [33]. Thus, knowledge derived from previous studies using static picture stimuli may be less transferrable to real-life situations.

Additionally, the extent to which neural VSS reactivity corresponds to individual ratings could help us gain insight into the etiology of related problematic behaviors. Problematic pornography use (PPU) is characterized by increasing time spent using VSS, failed attempts at reducing this time and disregarding other activities as well as negative consequences resulting from VSS use [34–36]. Since 2019 severe forms of this behavior can be clinically recognized in the International Classification of Disease (ICD-11; [6]) under the umbrella diagnosis Compulsive Sexual Behavior Disorder (CSBD). First studies have found altered striatal reactivity to VSS [29,37–39] and cues predicting VSS [40,41] in subjects reporting some form of compulsive sexual behavior compared to controls. The main drive behind PPU as precursory stage of CSBD is thought to be the positive reinforcement experienced by viewing VSS and carrying out related sexual behavior [42–44]. As PPU develops, VSS novelty and habituation mechanisms are discussed as important factors [45]. Thus, affected individuals have to seek out and use new VSS which best fit their individual preferences more frequently. Therefore, there is an increased need to identify highly preferred stimuli and respond to them. Accordingly, because striatal activity is associated with coding incentive salience, the neural response in NAcc and caudate should differentiate more strongly between VSS based on individual preference, the more PPU a subject reports. As the OFC is not necessarily implicated in these motivational processes, the same effect in this region is unlikely. In fact, Brand et al. [29] found, that the NAcc response to generally preferred VSS compared to non-preferred was positively correlated with self-reported PPU. This shows a connection of addictive behavior with stronger NAcc distinction between rewarding and non-rewarding stimuli. Whether an increased correspondence between individual ratings and neural activity with increasing PPU can be found in an individual correlational approach is unknown.

In sum, we aim first to examine the association between NAcc, caudate nucleus and OFC response to explicit, generally attractive VSS and individual ratings. We expect NAcc, caudate nucleus and OFC activity in response to viewing VSS to be positively correlated with the individual ratings of each VSS. We examine putamen activity as an exploratory analysis, but do not expect a significant correlation there. Second, we expect the strength of association between individual ratings and neural activity in NAcc to be increased the more PPU symptoms a subject reports. Although the caudate nucleus has not been investigated before, this increased rating/neural activity correspondence should also be found in the caudate nucleus because of its role in guiding goal-directed behavior. We examine the OFC as an exploratory analysis but do not expect a significant effect there.

2. Materials and methods

2.1. Participants

The final sample consisted of 72 healthy men (mean age = 25.56 y, SD = 4.45 y, range = 18–40). The data of 79 subjects were available for analysis. Of these, four had to be excluded due to technical difficulties, two because of image artefacts and one due to atypical neuroanatomy. Heterosexual and bisexual healthy men between 18 and 45 years were recruited through university e-mails and notice boards. Sexual orientation was measured using the Kinsey scale from 0 to 6 with 0 indicating absolutely heterosexual (mean sample score = 0.16, SD = .41, range = 0–2; [46]). Inclusion criteria were the absence of current somatic or mental diseases as well as of current psychotherapeutic or pharmacological treatment, no harmful use of alcohol or nicotine, no contra-indication for MRI and fluency in the German language. All participants had normal or corrected-to-normal vision and provided informed consent prior to any assessment. Subjects were informed that 'explicit pornographic material' would be shown and received 10 € per hour or course credit for their participation. The study was approved by a local ethics committee and was conducted in accordance with the 1964 declaration of Helsinki and its later amendments.

2.2. Procedure

Upon arrival at the lab, participants filled in the consent form and performed a short practice task of the paradigm described below outside of the scanner. As this sample consists of the control group of a larger study on VSS reactivity and acute psychosocial stress, the subjects then underwent a non-stressful control task [47] before MRI scanning. In the scanner, field map, anatomical and functional images were obtained. During the functional scan, the subjects performed the sexual incentive delay (SID) task. Two hours after the functional scan, participants were seated alone at a computer in a separate room where they filled out questionnaires and performed the rating task. The experimenter was not present in the room during this to ensure anonymity of the participant's answers.

2.3. Stimuli

An initial pool of 50 VSS film clips and 50 control film clips was

obtained from online video hosting platforms. All film clips were six seconds long and presented without sound. All VSS film clips showed at least one woman with a partner (manual stimulation, oral and vaginal intercourse) and contained no fetish-related material. Control clips depicted physiotherapeutic non-sexual massages. All film clips were rated on valence and sexual arousal by an independent sample of 58 non-homosexual men in a preliminary study. Both rating scales ranged from '1' (indicating 'very unpleasant' or 'not sexually arousing at all') to '9' (indicating 'very pleasant' or 'very sexually arousing'). Values from 5 to 9 were classified as high. The final stimuli set consisted of 21 VSS clips with high valence (M = 6.20, SD = 1.12) and high sexual arousal (M = 6.29, SD = 1.34) ratings from the preliminary study and 21 control clips with medium to high valence (M = 5.44, SD = 0.97) and low sexual arousal (M = 1.86, SD = 0.81) ratings from the preliminary study.

2.4. SID task

The procedure in the magnetic resonance image (MRI) scanner was adopted from the established monetary incentive delay task [48]. Instead of monetary rewards, we used film clips showing VSS or videos depicting non-sexual massages (control). The experiment was realized with the Presentation software package (Version 17.0, Presentation®, Neurobehavioral Systems Inc., Berkeley, USA).

The experimental task consisted of 63 trials with three conditions (21 x *VSS*, 21 x *Control,* 21 x *Nothing*) and lasted for about 20 min. Each trial consisted of an anticipation phase and a delivery phase. During the anticipation phase, one of three geometric figures (see Fig. 1) served as Cue_{VSS}, $Cue_{Control}$ or $Cue_{Nothing}$. Which figure served as which cue was balanced across participants.

In each trial, one of the three cues was presented for 4 s followed by a fixation cross for a variable (1–3 s) interstimulus interval. Next, a target (white square, 200 × 200 pixel) was presented for a minimum of 16 ms up to a maximum of 750 ms. The exact presentation time depended on an adaptive algorithm described below (resulted in overall target duration range of 44–624 ms). Subjects were instructed to press a reaction button every time the target was presented regardless of the cue presented before. Pressing the reaction button while the target was visible resulted in the win of a film clip if a Cue_{VSS} or a $Cue_{Control}$ was shown before. After target presentation, a fixation cross was presented for a variable (0–2 s) interstimulus interval, followed by the presentation of either a VSS clip (Cue_{VSS} & fast response), a control film clip ($Cue_{Control}$ & fast response) or a black screen (Cue_{VSS} & slow response, $Cue_{Control}$ & slow response, $Cue_{Nothing}$) for 6 s. After a variable (2–6 s) inter-trial interval, the next trial started.

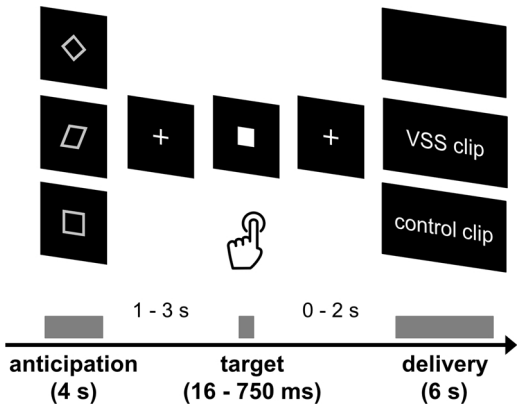

Fig. 1. Sexual incentive delay (SID) task. Subjects first saw one of three cues (geometric figures: Cue_{VSS}, $Cue_{Control}$, $Cue_{Nothing}$). After a variable delay, a target appeared for a short time. Subjects were instructed to push a button as soon as the target appeared. In trials that began with a Cue_{VSS} or a $Cue_{Control}$, fast reactions led to winning a VSS or control video clip. After the target had disappeared, the video clip or a black screen was displayed.

Before scanning, participants were informed, which figure served as which cue in the experiment and practiced the task. From this practice task, individual mean reaction times and standard deviations were obtained. During the main experiment, these were used to calculate target presentation times (win: $Mean_{RT} + 2 \times SD_{RT}$; loss: $Mean_{RT} - 2 \times SD_{RT}$). Presentation times were varied according to pre-scheduled reinforcement trials. Approximately 71 % of the *VSS* and *Control* trials (15 of 21 trials each) were scheduled for wins, while *Nothing* trials never resulted in a win. If subjects won unplanned or did not win in scheduled reinforcement trials, the target presentation time was corrected online (subtracting or adding 20 ms to the presentation time, respectively) to ensure reinforcement as planned in future trials. *VSS* and *Control* trials that did not result in wins or losses as planned were adaptively repeated in the next scheduled trials with the new duration of target presentation.

Participants were asked to rate their current level of sexual arousal on a 9-point Likert scale once before and once after the experimental task.

2.5. Individual ratings & questionnaire

Participants rated the stimuli set on valence and sexual arousal. All 21 control and 21 VSS film clips were presented, each followed by two nine-point Likert-type scales. Both scales ranged from '1' (indicating 'very unpleasant' or 'not sexually arousing at all') to '9' (indicating 'very pleasant' or 'very sexually arousing'). The resulting VSS ratings were mean-centered and included as parametric modulators in the functional MRI (fMRI) first level model.

Participants digitally filled out the German versions of the short Internet Addiction Test modified for cybersex, which can result in a total score of 12–60 from 12 items (s-IATsex; [49]). The items measure the experience of negative consequences and a loss of control regarding subjects' cybersex use as well as craving, social problems, preoccupation and mood regulation. The s-IATsex showed high internal consistency in this sample (α = .902). In addition to the questionnaire, subjects were asked to estimate from the last month, how much pornography they normally view in hours and minutes per day or per week. All answers were converted into hours per month during analysis.

All rating and questionnaire data were analyzed using SPSS 22 (Version 22.0, IBM Corp., Armonk, USA). Correlational analyses were run for VSS ratings, s-IATsex score and time spent viewing pornography. Control and VSS ratings as well as pre- and post- SID task sexual arousal ratings were compared with paired *t*-tests.

2.6. fMRI data acquisition and analysis

All images were acquired using a 3 T whole-body tomograph (Siemens Prisma) with a 64-channel head coil. The structural images consisted of 176 T1-weighted sagittal slices (MPRAGE; slice thickness 0.9 mm; FoV = 240 mm; TR = 1.58 s; TE = 2.3 s). For the functional images, a total of 632 images was acquired with a T2*-weighted gradient echo-planar imaging (EPI) with 36 slices covering the whole brain (voxel size = 3 × 3 × 3.5 mm; gap = 0.5 mm; descending slice acquisition; TR = 2 s; TE = 30 ms; flip angle = 75; FoV = 192 × 192 mm^2; matrix size = 64 × 64; GRAPPA = 2). The field of view was positioned automatically relative to the AC-PC line with an orientation of -30°.

Preprocessing, first and second level analysis was done using SPM 12 (Wellcome Department of Cognitive Neurology, London, UK) implemented in Matlab 2012 (The MathWorks Inc., Natick, Massachusetts, USA). For preprocessing, the anatomical image was coregistered to the Montreal Neurological Institute (MNI) space using the unified model implemented in SPM. All EPI images were realigned and unwarped using field maps, coregistered to the anatomical images, slice time

corrected, normalized to MNI standard space using the parameters from anatomy segmentation and smoothed with a Gaussian kernel at 6 mm FWHM (6 mm kernel as suggested by Sacchet and Knutson [50]). Functional data were screened for outlying volumes using a distribution free approach with thresholding for skewed data [51]. Each resulting outlying volume was later modeled within the general linear model as a regressor of no interest.

The experimental conditions were Cue_{VSS}, $Cue_{Control}$, $Cue_{Nothing}$, $Delivery_{VSS}$, $NoDelivery_{VSS}$, $Delivery_{Control}$, $NoDelivery_{Control}$, $NoDelivery_{Nothing}$ and *target*. Each condition was modelled as a regressor in the first level model. A parametric modulator containing the mean-centered ratings was added to the $Delivery_{VSS}$ regressor. This modulator modelled the $Delivery_{VSS}$ events weighted with the respective ratings. We set up two separate models, one with valence ratings and one with sexual arousal ratings as a modulator. All regressors were convolved with the canonical hemodynamic response function. Six movement parameters were entered as regressors of no interest. Additional regressors of no interest modelling the identified outlying volumes were entered as well. The time series was then filtered with a high pass filter (time constant = 128 s). One contrast of interest per model containing only the modulator was defined: '$Delivery_{VSS}$ x valence' and '$Delivery_{VSS}$ x sexual arousal'. On the group level, we performed two separate one-sample *t*-tests. One examining the correlation results for '$Delivery_{VSS}$ x valence' and one examining the correlation results for '$Delivery_{VSS}$ x sexual arousal'. Further, two linear regressions with the s-IATsex score as predictor and one contrast each as outcome were performed. One subject was identified as an outlier using Tukey Fences criterion (values were larger than 1.5 × interquartile range + third quartile; [52]) on the s-IATsex as well as both modular contrasts. This subject was excluded from the s-IATsex regressions leaving 71 subjects for these analyses.

Confirmatory region of interest (ROI) analyses on the voxel level were conducted using SPM's small volume correction with $p < 0.05$ (FWE). NAcc, caudate nucleus and OFC were chosen as ROIs because they have been previously reported in studies concerned with VSS processing [4]. The ROI masks for NAcc and caudate nucleus were taken from the Harvard Oxford Cortical Atlas, the OFC mask was created using MARINA [53]. Thus, there were three ROIs per hemisphere. For the exploratory putamen analysis, we also used the ROI masks from the Harvard Oxford Cortical Atlas (one per hemisphere). Explorative whole brain analyses on the voxel level were conducted with $p < 0.05$ family-wise-error (FWE) correction.

3. Results

3.1. Questionnaire & rating data

Descriptive data of ratings and questionnaire are shown in Table 1. VSS clips were rated significantly higher than control clips on both valence [$t(71) = 4.44$, $p < .001$] and sexual arousal [$t(71) = 30.20$, $p < .001$]. Subjects rated their current sexual arousal higher after the experiment compared to before [$t(70) = 7.17$, $p < .001$]. Mean s-IATsex score was moderate with a high variance. The s-IATsex scores were in line with those previously reported in comparable samples [29,54].

Mean valence and sexual arousal ratings of VSS were similar to those obtained in the preliminary rating study. In terms of correlations in the current study, VSS valence and sexual arousal scales were associated ($r = .659$, $p < .001$). The hours of pornography consumed per month correlated with the s-IATsex score ($r = .516$, $p < .001$). No other significant correlations between VSS ratings and individual reports were found (all *p*-values > .05).

Table 1
Descriptive statistics (mean, minimum, maximum, standard deviation) of rating and questionnaire data. Possible range of all ratings was 1-9. Possible range of s-IATsex was 12-60.

Variable	M	Min	Max	SD
VSS valence rating	6.38	2.14	8.67	1.17
VSS sexual arousal rating	6.65	2.14	8.62	1.15
Control valence rating	5.53	2.95	8.86	1.30
Control sexual arousal rating	1.99	1.00	5.00	.098
Sexual arousal rating pre- SID task	1.82	1	8	1.33
Sexual arousal rating post- SID task	3.49	1	9	1.88
s-IATsex	20.42	12	56	8.03
Pornography consumption (hours per month)	6.48	0	42	7.29

3.2. Hemodynamic responses

The BOLD response during VSS viewing was correlated with both valence and sexual arousal ratings in two ROIs. The higher the ratings of a VSS clip, the higher hemodynamic responses were in bilateral NAcc and bilateral caudate nucleus during viewing of the respective VSS clip. Hemodynamic responses in bilateral OFC correlated with valence ratings only. No significant correlation with sexual arousal ratings were observed in the OFC. The exploratory putamen analysis did not indicate any significant results for neither '$Delivery_{VSS}$ x valence' nor '$Delivery_{VSS}$ x sexual arousal'. For all results see Table 2 and Fig. 2. The explorative whole brain analyses revealed no significant results.

The regression results show that the s-IATsex score was positively associated with the correlation of BOLD response and sexual arousal ratings in left NAcc and bilateral caudate nucleus. There were no significant results with the valence ratings but a trend towards similar positive correlation effects was observed (see Table 3 and Fig. 3). The more problems with internet pornography use a subject reported, the stronger the association between hemodynamic responses in NAcc and caudate nucleus during VSS viewing and sexual arousal rating of the respective VSS clip. The explorative whole brain analyses revealed no significant results.

4. Discussion

This study addressed the relationship between individual stimuli ratings and the reward-related processing of visual sexual stimuli (VSS) in a large sample of healthy men. In the innovative sexual incentive delay (SID) task, VSS clips were used as rewards to be received through an action. Using pre-rated generally highly rewarding VSS, we tested whether individual VSS ratings are positively correlated with neural responses during viewing of the respective VSS. The higher a subject rated a VSS clip on sexual arousal or valence, the higher activity we

Table 2
Region of interest (ROI) results for the contrasts '$Delivery_{VSS}$ x valence' and '$Delivery_{VSS}$ x sexual arousal' (One Sample voxel level *t*-tests): Structure, side, coordinates (x,y,z, MNI space), cluster size (k), correlation (r) and statistics (FWE-corrected). Only results with $P_{corr} < 0.1$ are displayed.

Contrast	Structure	Side	x	y	z	k	T_{max}	r	P_{corr}
$Delivery_{VSS}$ x valence	NAcc	L	−8	10	−6	56	4.07	.43	.002**
		R	8	10	−4	45	3.30	.36	.014*
	Caudate	L	−8	6	6	352	3.58	.39	.029*
		R	8	8	6	218	3.35	.38	.038*
	OFC	L	−12	48	−2	583	4.08	.41	.016*
		R	8	52	−4	451	3.83	.43	.031*
	Putamen	R	20	10	2	161	3.27	.36	.088
$Delivery_{VSS}$ x sexual arousal	NAcc	L	−6	12	−6	59	3.37	.37	.014*
		R	8	12	−4	36	3.66	.41	.003**
	Caudate	L	−8	2	8	271	3.64	.39	.028*
		R	8	12	−2	233	4.18	.44	.006**
	OFC	L	−6	46	−6	423	3.70	.40	.052
	Putamen	L	−24	8	10	198	3.38	.37	.077

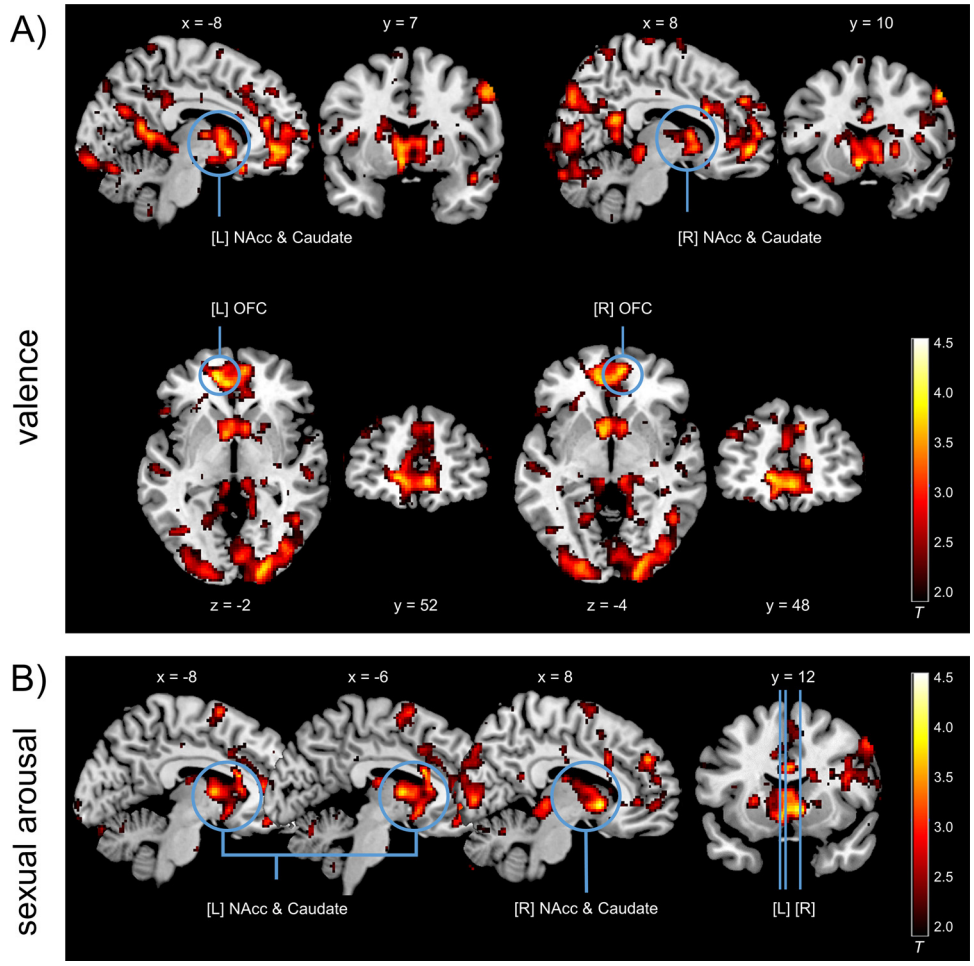

Fig. 2. Parametric modulation of ROI activity during viewing of visual sexual stimuli ($Delivery_{VSS}$) by subjective valence rating (A) and sexual arousal rating (B) of respective VSS on voxel level $p < .05$ (FWE-corrected within ROIs). Displayed t-values are thresholded at $t < 2$. A): NAcc and caudate nucleus (above) as well as OFC (below) activity is positively correlated with valence rating f viewed VSS. B): NAcc and caudate nucleus activity is positively correlated with sexual arousal rating of viewed VSS. Lines on the coronal slice on the right side indicate the sagittal slices depicted on the left.

Table 3
Region of Interest (ROI) results for the s-IATsex regression analyses with '*Delivery*$_{VSS}$ x valence' and '*Delivery*$_{VSS}$ x sexual arousal': s-IATsex score correlated with association between *Delivery*$_{VSS}$ and ratings with structure, side, coordinates (x,y,z, MNI space), cluster size (k), correlation (r) and statistics (FWE-corrected). Only results with $P_{corr} < 0.1$ are displayed.

Contrast	Structure	Side	x	y	z	k	T_{max}	r	P_{corr}
Delivery$_{VSS}$ x sexual arousal	NAcc	L	−8	12	−4	35	2.95	.33	.039*
		R	12	18	−6	18	2.71	.31	.062
	Caudate	L	−12	−6	18	368	4.44	.47	.003**
		R	14	−8	20	291	4.02	.43	.010**
Delivery$_{VSS}$ x valence	NAcc	L	−8	12	−6	23	2.72	.31	.061
	Caudate	L	−6	10	6	209	3.31	.37	.064

found in NAcc, caudate nucleus and OFC during VSS viewing. Additionally, the association between individual sexual arousal ratings and NAcc as well as caudate nucleus activity was stronger when subjects reported more symptoms of problematic pornography use (PPU) measured by the s-IATsex.

The correlation between bilateral NAcc and caudate nucleus responses to VSS and both individual VSS ratings of valence and sexual arousal indicates higher activity in these striatal regions when viewing VSS that the individual subject liked better. This suggests that components of the subjective VSS reward value could be coded in those regions.

Our results are consistent with past studies, where NAcc activity during VSS processing has been associated with subjective aspects of reward [30,31,55]. Where past research has shown NAcc activity to differ between pre-defined dichotomous VSS intensity or preference categories [30,32], our results extend this to a linear scaling of activity corresponding to individual preferences among generally highly arousing VSS. NAcc activity may code both incentive salience and valence of VSS as it has been found for other rewarding stimuli for salience [13,15] and valence [56,57]. While the caudate nucleus as part of the dorsomedial striatum has been found to be involved in VSS processing [37,58,59], linear scaling effects with individual ratings have not been reported before. Our results suggest that caudate nucleus activity aligns with individual liking, which might reflect the value of a VSS-directed action (e.g. sexual activity) based on the VSS preferences. It also indicates that the VSS may have been particularly motivationally salient in this study, possibly relating to the active nature of the task and because the stimuli were video clips. We did not find any correlation between VSS reactivity and individual ratings in the putamen (dorsolateral striatum). This was not surprising since putamen activity has consistently been found in connection with habitual stimulus-response-associations, but not sensitive to the outcome value, in previous studies [18,60].

We found that bilateral OFC reactivity to VSS correlated with individual reports of valence only. This result corroborates previous

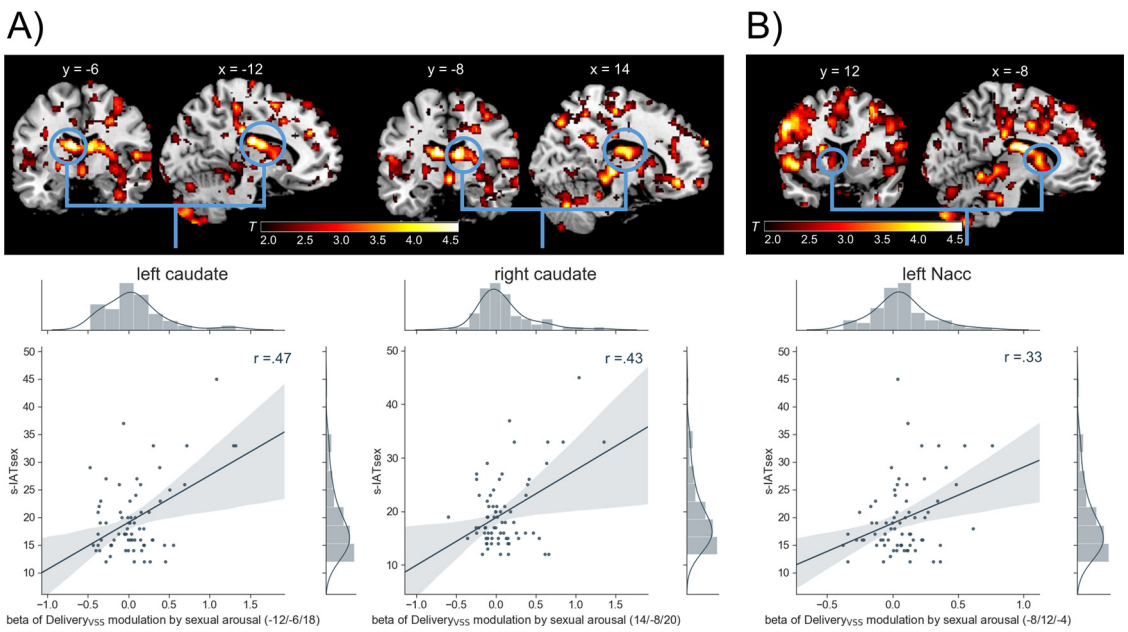

Fig. 3. Positive correlation between caudate nucleus (A) and NAcc (B) modulation by sexual arousal ratings during Delivery$_{VSS}$ and s-IATsex on voxel level $p < .05$ (FWE-corrected ROIs). Displayed *t*-values are thresholded at $t < 2$. Histograms and scatter plots of the contrast estimates at the respective peak voxel and s-IATsex score with linear regression line and 95 % confidence bands.

research on subjective reward values. OFC activity has been associated with valence ratings of many different rewards like attractive faces [15,61], food [22,23,62,63] and odors [25,62] and is thus seen as essential in coding hedonic experience [9]. Like the striatum, the OFC has been implicated in the coding of categorical VSS intensity before [30,32], but our results expand this to a linear relationship between ratings and neural activity among generally pleasant VSS. Here, we show an upward scaling of activity in the OFC when individuals viewed VSS that were very pleasant to them personally. Individual valence ratings probably mainly reflect the hedonic component of subjective VSS value while individual sexual arousal ratings would reflect the motivational component more. Therefore, these results might also serve as an indication that the OFC mainly codes the hedonic VSS value component, while the striatal regions code both incentive salience and hedonic pleasantness.

Taken together, the results fit well with the theory of a reward valuation system, which is independent of the type of reward. NAcc, caudate nucleus and OFC have been found in connection to subjective values for many different rewards [7,8] and we found a gradual increase in activity of these regions with increasing individual VSS liking. Most importantly, going further than previous findings on general and categorical VSS preference effects [28–30,32], we found the changes in activity to scale with VSS liking at an individual level using an individual correlational approach. This supports the hypothesized importance of individual preferences in VSS valuation.

These individual differences in preference coding might represent a mechanism that mediates addictive VSS use experienced by some individuals. We not only found an association of NAcc and caudate activity with sexual arousal ratings during VSS viewing but the strength of this association was greater when the subject reported more PPU. The result supports the hypothesis, that incentive value responses in NAcc and caudate differentiate more strongly between differently preferred stimuli, the more a subject experiences PPU. This extends past studies, where PPU has been linked to a higher striatal response to VSS as compared to a control or non-preferred condition [29,38]. One study, also using an SID task, found increased NAcc activity associated with increased PPU during the anticipation phase only [41]. Our results indicate that a similar effect, i.e. altered incentive salience processing associated with PPU, can also be found in the delivery phase, but only if individual preference is taken into account. The increasing differentiation of incentive value signals in the NAcc could reflect an increased need for seeking and identifying preferred VSS during addiction development. As long as longitudinal studies are pending, this, however, remains speculation. The caudate result suggests that individual VSS preference could have a much stronger influence on the likelihood of engaging in sexual activity when subjects experience more PPU.

Given these results can be replicated, they may have important clinical implications. Increased differentiation of incentive value signals might be connected to an increase in time spent searching for highly stimulating material, which later leads to issues in personal or professional life and suffering because of this behavior. Regarding possible tolerance development, it would be interesting if the differences between values got larger over time, requiring the search for and use of more and more highly preferred VSS to obtain the same reward. So far, however, it remains unknown whether this effect is precursor or result of the addictive behavior.

A number of strengths and methodological improvements may underscore the value of our findings. The standardized pre-rated stimuli in conjunction with individual ratings enabled us to consider personal preferences while retaining high internal validity. The observation that regions differentiated according to individual preference, although the stimuli were highly attractive to most subjects, suggests a rather fine resolution of valuation. Furthermore, the ratings successfully predicted brain reactivity during the task although they were obtained afterwards, suggesting adequate reliability of these individual reports. We used the SID task to include active processes in the design. The original version with monetary rewards is already widely used [11,48] and versions with sexual stimuli have recently started to gain popularity [31,41]. This task is probably much closer to real-life search and use behavior than passive viewing tasks used in previous studies. We deliberately chose to use film clips instead of pictures in the task to further increase external validity. This task combined with these stimuli probably aided in eliciting robust preference-associated neural responses, which we could then further examine in association with the questionnaire data. Finally, we were able to collect and analyze a very large sample, which is still rare in fMRI research. Of course, there are some questions, we could not address in this study. We investigated an exclusively male non-homosexual sample. To understand VSS reward

valuation and PPU better in the future, the effect of individual preference should also be examined in a more gender- and sexual orientation-diverse sample. Longitudinal studies should also be considered to investigate the underlying causal explanation for the differences in value processing associated with addiction symptoms.

In conclusion, the results demonstrate that the neural responses in NAcc, caudate, and OFC closely code individual preferences among highly rewarding VSS as never shown before. The association between PPU and the correspondence between individual ratings and neural activity might allow new insights in the neurobiological mechanisms underlying the development of a CSBD.

CRediT authorship contribution statement

Sanja Klein: Investigation, Formal analysis, Writing - original draft, Visualization. **Onno Kruse:** Conceptualization, Methodology, Investigation, Writing - review & editing. **Charlotte Markert:** Investigation, Writing - review & editing. **Isabell Tapia León:** Writing - review & editing. **Jana Strahler:** Conceptualization, Methodology, Investigation, Writing - review & editing. **Rudolf Stark:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgments

This study was funded by the German research association (DFG STA 475/16-1).

References

[1] W. Schultz, Neuronal reward and decision signals: from theories to data, Physiol. Rev. 95 (2015) 853–951, https://doi.org/10.1152/physrev.00023.2014.
[2] J. Zhang, K.C. Berridge, A.J. Tindell, K.S. Smith, J.W. Aldridge, A neural computational model of incentive salience, PLoS Comput. Biol. 5 (2009), https://doi.org/10.1371/journal.pcbi.1000437.
[3] J.R. Georgiadis, M.L. Kringelbach, The human sexual response cycle: brain imaging evidence linking sex to other pleasures, Prog. Neurobiol. 98 (2012) 49–81, https://doi.org/10.1016/j.pneurobio.2012.05.004.
[4] S. Stoléru, V. Fonteille, C. Cornélis, C. Joyal, V. Moulier, Functional neuroimaging studies of sexual arousal and orgasm in healthy men and women: a review and meta-analysis, Neurosci. Biobehav. Rev. 36 (2012) 1481–1509, https://doi.org/10.1016/j.neubiorev.2012.03.006.
[5] J.B. Grubbs, P.J. Wright, A.L. Braden, J.A. Wilt, S.W. Kraus, Internet pornography use and sexual motivation: a systematic review and integration, Ann. Int. Comm. Assoc. 43 (2019) 117–155, https://doi.org/10.1080/23808985.2019.1584045.
[6] World Health Organization, ICD-11 for Mortality and Morbidity Statistics, (2019) https://icd.who.int/en/.
[7] O. Bartra, J.T. McGuire, J.W. Kable, The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value, NeuroImage 76 (2013) 412–427, https://doi.org/10.1016/j.neuroimage.2013.02.063.
[8] S. Kühn, J. Gallinat, The neural correlates of subjective pleasantness, NeuroImage 61 (2012) 289–294, https://doi.org/10.1016/j.neuroimage.2012.02.065.
[9] J. Peters, C. Büchel, Neural representations of subjective reward value, Behav. Brain Res. 213 (2010) 135–141, https://doi.org/10.1016/j.bbr.2010.04.031.
[10] S.N. Haber, B. Knutson, The reward circuit: linking primate anatomy and human imaging, Neuropsychopharmacology 35 (2010) 4–26, https://doi.org/10.1038/npp.2009.129.
[11] R.P. Wilson, M. Colizzi, M.G. Bossong, P. Allen, M. Kempton, S. Bhattacharyya, The neural substrate of reward anticipation in health: a meta-analysis of fMRI findings in the monetary incentive delay task, Neuropsychol. Rev. 28 (2018) 496–506, https://doi.org/10.1007/s11065-018-9385-5.
[12] S. Oldham, C. Murawski, A. Fornito, G. Youssef, M. Yücel, V. Lorenzetti, The anticipation and outcome phases of reward and loss processing: a neuroimaging meta-analysis of the monetary incentive delay task, Hum. Brain Mapp. 39 (2018) 3398–3418, https://doi.org/10.1002/hbm.24184.
[13] B. Knutson, C.M. Adams, G.W. Fong, D. Hommer, Anticipation of increasing monetary reward selectively recruits nucleus accumbens, J. Neurosci. 21 (2001) RC159, https://doi.org/10.1523/JNEUROSCI.21-16-j0002.2001.
[14] M. Lebreton, S. Jorge, V. Michel, B. Thirion, M. Pessiglione, An automatic valuation system in the human brain: evidence from functional neuroimaging, Neuron 64 (2009) 431–439, https://doi.org/10.1016/j.neuron.2009.09.040.
[15] J. Cloutier, T.F. Heatherton, P.J. Whalen, W.M. Kelley, Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness, J. Cogn. Neurosci. 20 (2008) 941–951, https://doi.org/10.1162/jocn.2008.20062.
[16] E.T. Rolls, F. Grabenhorst, B.A. Parris, Warm pleasant feelings in the brain, NeuroImage 41 (2008) 1504–1513, https://doi.org/10.1016/j.neuroimage.2008.03.005.
[17] B.J. Everitt, T.W. Robbins, From the ventral to the dorsal striatum: devolving views of their roles in drug addiction, Neurosci. Biobehav. Rev. 37 (2013) 1946–1954, https://doi.org/10.1016/j.neubiorev.2013.02.010.
[18] B.W. Balleine, J.P. O'Doherty, Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action, Neuropsychopharmacology 35 (2010) 48–69, https://doi.org/10.1038/npp.2009.131.
[19] S. Luo, G. Ainslie, L. Giragosian, J.R. Monterosso, Behavioral and neural evidence of incentive bias for immediate rewards relative to preference-matched delayed rewards, J. Neurosci. 29 (2009) 14820–14827, https://doi.org/10.1523/JNEUROSCI.4261-09.2009.
[20] T.M. Le, S. Zhang, S. Zhornitsky, W. Wang, C.-S.R. Li, Neural correlates of reward-directed action and inhibition of action, Cortex 123 (2020) 42–56, https://doi.org/10.1016/j.cortex.2019.10.007.
[21] A. Pine, B. Seymour, J.P. Roiser, P. Bossaerts, K.J. Friston, H.V. Curran, R.J. Dolan, Encoding of marginal utility across time in the human brain, J. Neurosci. 29 (2009) 9575–9581, https://doi.org/10.1523/JNEUROSCI.1126-09.2009.
[22] T.A. Hare, C.F. Camerer, A. Rangel, Self-control in decision-making involves modulation of the vmPFC valuation system, Science 324 (2009) 646–648, https://doi.org/10.1126/science.1168450.
[23] H. Plassmann, J.P. O'Doherty, B. Shiv, A. Rangel, Marketing actions can modulate neural representations of experienced pleasantness, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 1050–1054, https://doi.org/10.1073/pnas.0706929105.
[24] E.T. Rolls, C. McCabe, Enhanced affective brain representations of chocolate in cravers vs. non-cravers, Eur. J. Neurosci. 26 (2007) 1067–1076, https://doi.org/10.1111/j.1460-9568.2007.05724.x.
[25] J.-P. Royet, J. Plailly, C. Delon-Martin, D.A. Kareken, C. Segebarth, fMRI of emotional responses to odors: influence of hedonic valence and judgment, handedness, and gender, NeuroImage 20 (2003) 713–728, https://doi.org/10.1016/S1053-8119(03)00388-4.
[26] H. Kim, S. Shimojo, J.P. O'Doherty, Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex, Cereb. Cortex 21 (2011) 769–776, https://doi.org/10.1093/cercor/bhq145.
[27] T.B. Poeppl, B. Langguth, A.R. Laird, S.B. Eickhoff, The functional neuroanatomy of male psychosexual and physiosexual arousal: a quantitative meta-analysis, Hum. Brain Mapp. 35 (2014) 1404–1421, https://doi.org/10.1002/hbm.22262.
[28] J. Ponseti, H.A. Bosinski, S. Wolff, M. Peller, O. Jansen, H.M. Mehdorn, C. Büchel, H.R. Siebner, A functional endophenotype for sexual orientation in humans, NeuroImage 33 (2006) 825–833, https://doi.org/10.1016/j.neuroimage.2006.08.002.
[29] M. Brand, J. Snagowski, C. Laier, S. Maderwald, Ventral striatum activity when watching preferred pornographic pictures is correlated with symptoms of internet pornography addiction, NeuroImage 129 (2016) 224–232, https://doi.org/10.1016/j.neuroimage.2016.01.033.
[30] G. Sescousse, J. Redouté, J.-C. Dreher, The architecture of reward value coding in the human orbitofrontal cortex, J. Neurosci. 30 (2010) 13095–13104, https://doi.org/10.1523/JNEUROSCI.3501-10.2010.
[31] G. Sescousse, G. Barbalat, P. Domenech, J.-C. Dreher, Imbalance in the sensitivity to different types of rewards in pathological gambling, Brain 136 (2013) 2527–2538, https://doi.org/10.1093/brain/awt126.
[32] M. Walter, F. Bermpohl, H. Mouras, K. Schiltz, C. Tempelmann, M. Rotte, H.J. Heinze, B. Bogerts, G. Northoff, Distinguishing specific sexual and general emotional effects in fMRI-subcortical and cortical arousal during erotic picture viewing, NeuroImage 40 (2008) 1482–1494, https://doi.org/10.1016/j.neuroimage.2008.01.040.
[33] I. Solano, N.R. Eaton, K.D. O'Leary, Pornography consumption, modality and function in a large internet sample, J. Sex Res. 57 (2020) 92–103, https://doi.org/10.1080/00224499.2018.1532488.
[34] A. Kor, Y. Fogel, R.C. Reid, M.N. Potenza, Should hypersexual disorder be classified as an addiction? Sex Addict. Compulsivity 20 (2013), https://doi.org/10.1080/10720162.2013.768132.
[35] R. Stark, T. Klucken, M.N. Potenza, M. Brand, J. Strahler, A current understanding of the behavioral neuroscience of compulsive sexual behavior disorder and problematic pornography use, Curr. Behav. Neurosci. Rep. 5 (2018) 218–231, https://doi.org/10.1007/s40473-018-0162-9.
[36] M.P. Kafka, Hypersexual disorder: a proposed diagnosis for DSM-V, Arch. Sex. Behav. 39 (2010) 377–400, https://doi.org/10.1007/s10508-009-9574-7.
[37] J.-W. Seok, J.-H. Sohn, Neural substrates of sexual desire in individuals with problematic hypersexual behavior, Front. Behav. Neurosci. 9 (2015) 321, https://doi.org/10.3389/fnbeh.2015.00321.
[38] V. Voon, T.B. Mole, P. Banca, L. Porter, L.S. Morris, S. Mitchell, T.R. Lapa, J. Karr, N.A. Harrison, M.N. Potenza, M. Irvine, Neural correlates of sexual cue reactivity in individuals with and without compulsive sexual behaviours, PLoS One 9 (2014) e102419, , https://doi.org/10.1371/journal.pone.0102419.
[39] S. Kühn, J. Gallinat, Brain structure and functional connectivity associated with pornography consumption: the brain on porn, JAMA Psychiatry 71 (2014) 827–834, https://doi.org/10.1001/jamapsychiatry.2014.93.
[40] T. Klucken, S. Wehrum-Osinsky, J. Schweckendiek, O. Kruse, R. Stark, Altered appetitive conditioning and neural connectivity in subjects with compulsive sexual behavior, J. Sex. Med. 13 (2016) 627–636, https://doi.org/10.1016/j.jsxm.2016.01.013.
[41] M. Gola, M. Wordecha, G. Sescousse, M. Lew-Starowicz, B. Kossowski, M. Wypych, S. Makeig, M.N. Potenza, A. Marchewka, Can pornography be addictive? An fMRI study of men seeking treatment for problematic pornography use,

Neuropsychopharmacology 42 (2017) 2021–2031, https://doi.org/10.1038/npp.2017.78.
[42] K.S. Young, Internet sex addiction, Am. Behav. Sci. 52 (2008) 21–37, https://doi.org/10.1177/0002764208321339.
[43] C. Laier, M. Brand, Empirical evidence and theoretical considerations on factors contributing to cybersex addiction from a cognitive-behavioral view, Sex Addict. Compulsivity 21 (2014) 305–321, https://doi.org/10.1080/10720162.2014.970722.
[44] M. Brand, E. Wegmann, R. Stark, A. Müller, K. Wölfling, T.W. Robbins, M.N. Potenza, The interaction of Person-Affect-Cognition-Execution (I-PACE) model for addictive behaviors: update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors, Neurosci. Biobehav. Rev. 104 (2019) 1–10, https://doi.org/10.1016/j.neubiorev.2019.06.032.
[45] P. Banca, L.S. Morris, S. Mitchell, N.A. Harrison, M.N. Potenza, V. Voon, Novelty, conditioning and attentional bias to sexual rewards, J. Psychiatr. Res. 72 (2016) 91–101, https://doi.org/10.1016/j.jpsychires.2015.10.017.
[46] A.C. Kinsey, W. Pomeroy, C.E. Martin, Sexual Behavior in the Human Male, (1948).
[47] S. Het, N. Rohleder, D. Schoofs, C. Kirschbaum, O.T. Wolf, Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test', Psychoneuroendocrinology 34 (2009) 1075–1086, https://doi.org/10.1016/j.psyneuen.2009.02.008.
[48] B. Knutson, A. Westdorp, E. Kaiser, D. Hommer, FMRI visualization of brain activity during a monetary incentive delay task, NeuroImage 12 (2000) 20–27, https://doi.org/10.1006/nimg.2000.0593.
[49] C. Laier, M. Pawlikowski, J. Pekal, F.P. Schulte, M. Brand, Cybersex addiction: experienced sexual arousal when watching pornography and not real-life sexual contacts makes the difference, J. Behav. Addict. 2 (2013) 100–107, https://doi.org/10.1556/JBA.2.2013.002.
[50] M.D. Sacchet, B. Knutson, Spatial smoothing systematically biases the localization of reward-related brain activity, NeuroImage 66 (2013), https://doi.org/10.1016/j.neuroimage.2012.10.056.
[51] J. Schweckendiek, T. Klucken, C.J. Merz, S. Kagerer, B. Walter, D. Vaitl, R. Stark, Learning to like disgust: neuronal correlates of counterconditioning, Front. Hum. Neurosci. 7 (2013) 346, https://doi.org/10.3389/fnhum.2013.00346.
[52] J.W. Tukey, Exploratory Data Analysis, Addison-Wesley, Reading, Mass, 1997.
[53] B. Walter, C. Blecker, P. Kirsch, G. Sammer, A. Schienle, R. Stark, D. Vaitl, MARINA: an easy to use tool for the creation of MAsks for Region of INterest Analyses [abstract], Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19-22, 2003 New York, NY.
[54] J. Engel, A. Kessler, M. Veit, C. Sinke, I. Heitland, J. Kneer, U. Hartmann, T.H.C. Krüger, Hypersexual behavior in a large online sample: individual characteristics and signs of coercive sexual behavior, J. Behav. Addict. 8 (2019) 213–222, https://doi.org/10.1556/2006.8.2019.16.
[55] D. Sabatinelli, M.M. Bradley, P.J. Lang, V.D. Costa, F. Versace, Pleasure rather than salience activates human nucleus accumbens and medial prefrontal cortex, J. Neurophysiol. 98 (2007) 1374–1379, https://doi.org/10.1152/jn.00230.2007.
[56] A.B.M. Gerdes, M.J. Wieser, A. Mühlberger, P. Weyers, G.W. Alpers, M.M. Plichta, F. Breuer, P. Pauli, Brain activations to emotional pictures are differentially associated with valence and arousal ratings, Front. Hum. Neurosci. 4 (2010) 175, https://doi.org/10.3389/fnhum.2010.00175.
[57] J.C. Cooper, B. Knutson, Valence and salience contribute to nucleus accumbens activation, NeuroImage 39 (2008) 538–547, https://doi.org/10.1016/j.neuroimage.2007.08.009.
[58] H. Graf, M. Wiegers, C.D. Metzger, M. Walter, G. Grön, B. Abler, Erotic stimulus processing under amisulpride and reboxetine: a placebo-controlled fMRI study in healthy subjects, Int. J. Neuropsychopharmacol. 18 (2014), https://doi.org/10.1093/ijnp/pyu004.
[59] C.D. Metzger, U. Eckert, J. Steiner, A. Sartorius, J.E. Buchmann, J. Stadler, C. Tempelmann, O. Speck, B. Bogerts, B. Abler, M. Walter, High field fMRI reveals thalamocortical integration of segregated cognitive and emotional processing in mediodorsal and intralaminar thalamic nuclei, Front. Neuroanat. 4 (2010) 138, https://doi.org/10.3389/fnana.2010.00138.
[60] A.O. Ceceli, E.M. Tricomi, Habits and goals: a motivational perspective on action control, Curr. Opin. Behav. Sci. 20 (2018) 110–116, https://doi.org/10.1016/j.cobeha.2017.12.005.
[61] T.K. Pegors, J.W. Kable, A. Chatterjee, R.A. Epstein, Common and unique representations in pFC for face and place attractiveness, J. Cogn. Neurosci. 27 (2015) 959–973, https://doi.org/10.1162/jocn_a_00777.
[62] T. Jiang, R. Soussignan, B. Schaal, J.-P. Royet, Reward for food odors: an fMRI study of liking and wanting as a function of metabolic state and BMI, Soc. Cogn. Affect. Neurosci. 10 (2015) 561–568, https://doi.org/10.1093/scan/nsu086.
[63] W.K. Simmons, K.M. Rapuano, J.E. Ingeholm, J. Avery, S. Kallman, K.D. Hall, A. Martin, The ventral pallidum and orbitofrontal cortex support food pleasantness inferences, Brain Struct. Funct. 219 (2014) 473–483, https://doi.org/10.1007/s00429-013-0511-0.

Publication 2: Cross-paradigm integration shows a common neural basis for aversive and appetitive conditioning

NeuroImage 263 (2022) 119594

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Cross-paradigm integration shows a common neural basis for aversive and appetitive conditioning

Sanja Klein[a,b,d,*], Onno Kruse[a,b], Isabell Tapia León[b,c], Lukas Van Oudenhove[e,f,g], Sophie R. van 't Hof[h], Tim Klucken[c], Tor D. Wager[g], Rudolf Stark[a,b,d]

[a] Department of Psychotherapy and Systems Neuroscience, Justus Liebig University, Giessen 35394, Germany
[b] Bender Institute for Neuroimaging (BION), Justus Liebig University, Giessen 35394, Germany
[c] Clinical Psychology and Psychotherapy, University Siegen, Siegen 57076, Germany
[d] Center of Mind, Brain and Behavior, Universities of Marburg and Giessen, Marburg 35032, Germany
[e] Department of Chronic Diseases and Metabolism (CHROMETA), Laboratory for Brain-Gut Axis Studies (LaBGAS), Translational Research Centre for Gastrointestinal Disorders TARGID, KU Leuven, Leuven, Belgium
[f] Leuven Brain Institute, KU Leuven, Leuven, Belgium
[g] Department of Psychological and Brain Sciences, Cognitive and Affective Neuroscience Lab, Dartmouth College, Hanover, NH, USA
[h] Department of Psychiatry, Amsterdam University Medical Centers, Amsterdam 1105 AZ, The Netherlands

ARTICLE INFO

Keywords:
fMRI
Conditioning
Learning
Fear
Reward
Striatum

ABSTRACT

Sharing imaging data and comparing them across different psychological tasks is becoming increasingly possible as the open science movement advances. Such cross-paradigm integration has the potential to identify commonalities in findings that neighboring areas of study thought to be paradigm-specific. However, even the integration of research from closely related paradigms, such as aversive and appetitive classical conditioning is rare – even though qualitative comparisons already hint at how similar the 'fear network' and 'reward network' may be. We aimed to validate these theories by taking a multivariate approach to assess commonalities across paradigms empirically. Specifically, we quantified the similarity of an aversive conditioning pattern derived from meta-analysis to appetitive conditioning fMRI data. We tested pattern expression in three independent appetitive conditioning studies with 29, 76 and 38 participants each. During fMRI scanning, participants in each cohorts performed an appetitive conditioning task in which a CS+ was repeatedly rewarded with money and a CS- was never rewarded. The aversive pattern was highly similar to appetitive CS+ > CS- contrast maps across samples and variations of the appetitive conditioning paradigms. Moreover, the pattern distinguished the CS+ from the CS- with above-chance accuracy in every sample. These findings provide robust empirical evidence for an underlying neural system common to appetitive and aversive learning. We believe that this approach provides a way to empirically integrate the steadily growing body of fMRI findings across paradigms.

1. Introduction

Comparing paradigms and results across research areas is necessary to advance knowledge in basic and translational neuroscience. But even very closely related areas of research are often studied in parallel, accumulating data with little cross-fertilization between areas and their respective paradigms. Two such areas are the neural basis of fear learning and reward learning - conceptualized as aversive and appetitive conditioning, respectively. When these intrinsically adaptive learning processes become excessive, they can become the basis for psychological disorders such as anxiety, depression and addiction (Duits et al., 2015; Martin-Soelch et al., 2007). This conceptual distinction is reflected in the Research Domain Criteria (RDoC) framework, with 'fear learning' and 'reward learning' belonging to the separate domains of negative and positive valence systems (Insel, 2014). However, possible common underlying or interacting factors in these disorders (Destoop et al., 2019; Liverant et al., 2014; Xie et al., 2021) can be easily overlooked when we only examine these domains separately. Thus, shedding light on commonalities regarding their basic neural processes is essential going forward. Some efforts have been made to translate neuroimaging evidence from aversive to appetitive conditioning paradigms, but limited to qualitative comparisons and narrative reviews (e.g. Brooks and Berns 2013, Moscarello and LeDoux 2013, Stefanova et al. 2020). Only very recently, a meta-analysis on prediction errors also included a look at appetitive and aversive stimuli at a global level (Corlett et al., 2022). So far, no empirical integration of neuroimaging data from specifically aversive and appetitive conditioning studies has been attempted, although a lot of imaging data – especially on aversive conditioning – already exists and

* Corresponding author at: University of Giessen, Otto-Behaghel-Str. 10 H, Giessen 35394, Germany.
E-mail address: sanja.klein.psychol@gmail.com (S. Klein).

https://doi.org/10.1016/j.neuroimage.2022.119594.
Received 3 March 2022; Received in revised form 22 July 2022; Accepted 25 August 2022
Available online 28 August 2022.

Table 1
A detailed overview of regions reported in meta-analyses as well as theoretical models of aversive and appetitive learning. Common regions between aversive and appetitive CS+ are amygdala, NAcc, caudate nucleus, putamen, insula and thalamus.

	Aversive conditioning	Appetitive conditioning
Theoretical models	amygdala, mPFC, hippocampus Tovote et al. (2015)	amygdala, OFC, dACC, vACC, NAcc, caudate nucleus, putamen Martin-Soelch et al. (2007)
	amygdala, mPFC, dmPFC Herry and Johansen (2014)	amygdala, NAcc Averbeck and Costa (2017)
Empirical meta-analytic evidence	dACC, thalamus, anterior insular cortex, amygdala, OFC, putamen, midbrain/substantia nigra Etkin and Wager (2007)	amygdala, NAcc, caudate nucleus, putamen, midbrain, thalamus, frontal operculum, insula Chase et al. (2015)
	amygdala (smaller effects, only in uninstructed studies), anterior insula, putamen, caudate nucleus, dmPFC, dACC, preSMA, thalamus, pallidum Mechias et al. (2010)	
	anterior insular cortex, NAcc, caudate, SMA/preSMA, dlPFC, precuneus, cerebellum Fullana et al. (2016)	

Abbreviations: Nucleus Accumbens (NAcc), prefrontal cortex (PFC), medial PFC (mPFC), dorsomedial PFC (dmPFC), dorsolateral PFC (dlPFC), supplementar motor area (SMA), orbitofrontal cortex (OFC), dorsal/ventral anterior cingulate cortex (dACC/vACC)

qualitatively, activation patterns seem similar. Therefore, our general aims were, first, to attempt the empirical integration of data across the paradigms of aversive and appetitive conditioning. Second, we wanted to demonstrate the feasibility of integrating findings from these two paradigms in order to enable further research across a multitude of other paradigms of varying similarity.

The differential aversive or appetitive conditioning paradigms that are employed in fMRI research in humans are highly alike. An initially neutral stimulus becomes a conditioned stimulus (CS+) after repeated pairing with an aversive or appetitive unconditioned stimulus (UCS, e.g. electric shock or money). A second stimulus (CS-) is never paired with a UCS (Mackintosh, 1975). On the one hand these are striking similarities, on the other hand reward and fear seem diametrically opposed leading to separate investigations into brain regions constituting a fear network or a reward network. The neural correlates of aversive conditioning have been researched extensively in human neuroimaging, which has led to a large body of fMRI results on the topic as well as meta-analyses (for reviews see Etkin and Wager 2007, Fullana et al. 2016, Mechias et al. 2010, Sehlmeyer et al. 2009). In parallel, fMRI studies on appetitive conditioning have begun to accumulate (for reviews see Averbeck and Costa 2017, Chase et al. 2015, Martin-Soelch et al. 2007). It has become increasingly apparent that the findings from aversive and appetitive conditioning are qualitatively similar. The same regions often emerge from separate meta-analyses of responses to a CS+ compared to a CS- in aversive (Etkin and Wager, 2007; Fullana et al., 2016; Mechias et al., 2010) and appetitive (Chase et al., 2015) conditioning, see Table 1 for details. Seminal theoretical models of aversive conditioning focus mainly on the amygdala (Herry and Johansen, 2014; Tovote et al., 2015) while appetitive conditioning models also include striatal regions such as the Nucleus Accumbens (NAcc; Averbeck and Costa 2017, Martin-Soelch et al. 2007). In summary, the amygdala, NAcc, caudate nucleus, putamen, insula and thalamus seem to be involved in both aversive and appetitive learning, based on qualitative comparison of empirical data as well as theoretical models (see Table 1). The cerebellum has been reported in the most recent meta-analysis of aversive learning (Fullana et al., 2016) and since then in another aversive conditioning study in humans (Ernst et al., 2019). This region may be crucial for many different types of outcome prediction (Popa and Ebner, 2018) and has been shown associated with appetitive prediction in animal data (Heffley and Hull, 2019), so cerebellar activity might be another possible commonality between human aversive and appetitive learning. Based on these apparently overlapping regions, it is assumed that the concepts 'fear network' and 'reward network' share mesolimbic dopamine pathways and thus may share a common basis in an anticipatory motivational system related to learning in general (Menon and Uddin, 2010; Moscarello and LeDoux, 2013; Seeley et al., 2007; Stefanova et al., 2020). However, these assumptions are mostly based on qualitative literature reviews. Only few neuroimaging studies have systematically compared aversive and appetitive learning in the same experiment and even then mostly focused on differences instead of similarities (e.g. Breiter et al. 2001, Carter et al. 2009, Lake et al. 2019, Sankar et al. 2019). While elucidating the differences between these mechanisms remains important, quantifying cross-paradigm similarities might provide an even greater opportunity.

In this paper, we adopt a multivariate analysis approach to quantitatively integrate previously published evidence across paradigms and samples in order to better understand the commonalities of aversive and appetitive processes. With the help of machine learning classification algorithms, we can test whether whole-brain patterns of activation are present in a dataset and whether they distinguish between conditions (Weaverdyck et al., 2020; Woo et al., 2017). Multivariate approaches have already been used to great success in finding and validating whole brain response patterns associated with cognitive and affective states, e.g. the experience of pain (Wager et al., 2013), emotions (Kragel and LaBar, 2014; Saarimäki et al., 2016) or perceiving sexual pictures (van 't Hof et al., 2021; for a review on neural signatures see Kragel et al. (2018) based on data from the same kind of paradigm. Here, instead of developing an activation model from similar paradigms, we apply an already existing meta-analytic response pattern from one paradigm (aversive conditioning) to data from a similar paradigm (appetitive conditioning) to empirically identify activation commonalities. Using a meta-analytical pattern instead of training a new aversive conditioning pattern enables us to investigate similarity of our current appetitive conditioning data with the summarized data of numerous past aversive conditioning studies, gathered over many years of research.

In this study, we aim to identify commonalities of a differential activation pattern related to aversive conditioning, based on the meta-analysis by Fullana et al. (2016), with activation patterns in appetitive conditioning paradigms. In order to assess generalizability of the similarities, we carry out the same tests in three independent appetitive conditioning datasets with varying features regarding sample characteristics and procedural details (Kruse et al., 2018, 2020; Tapia León et al., 2019). First, we expect that the brain activation difference between aversive CS+ (avCS+) and aversive CS- (avCS-) will be similar to the activation difference between appetitive CS+ (appCS+) and appetitive CS- (appCS-), measured by a pattern expression score. We expect

this for differential activation over the whole brain as well as for a priori anatomical regions of interest (ROIs: NAcc, caudate nucleus, putamen, amygdala, thalamus, insula, cerebellum), which have been implicated in both forms of learning empirically and theoretically but may have traditionally been associated with one paradigm more than the other. Second, we hypothesize that the separate appCS+ and appCS- activation data will differ in their similarity to the avCS+ > avCS- pattern. We expect to accurately discriminate whether a pattern expression score stems from whole brain appCS+ or appCS- data based on the score's size via forced-choice classification. With these analyses, we aim to provide empirical evidence for the neural commonalities of aversive and appetitive conditioning at whole brain and region level.

2. Materials and methods

2.1. Sample descriptions

We used three previously published datasets on appetitive conditioning. All studies were approved by the local ethics committee and were conducted in accordance with the 1964 declaration of Helsinki and its later amendments. Participants gave written informed consent and received 10 € per hour or course credit for their participation plus monetary gains from the tasks.

2.1.1. Active learning/homogeneous sample

The Active Learning/Homogeneous Sample included only male subjects and a between-person acute stress condition (Kruse et al., 2018, see also Kruse et al. (2017). For our analysis, we included only the no-stress control group (n = 29, control group from Kruse et al. (2018) for this analysis. The mean age was M = 23.83 (SD = 2.80). Because this sample was the control group in a strictly timed stress experiment, the overall procedure was more rigorously controlled and standardized than in the other two samples.

2.1.2. Active learning/heterogeneous sample

The Active Learning/Heterogeneous Sample was larger (n = 76, Kruse et al. 2020) and included 36 men and 40 women with a mean age of M = 23.76 (SD = 3.73).

2.1.3. Passive learning/heterogeneous sample

The Passive Learning/Heterogeneous Sample (n = 38, Tapia León et al. 2019) also included men as well as women (22 men, 16 women) with a mean age of M = 23.50 (SD = 3.54).

2.2. Conditioning paradigms

2.2.1. Active learning/homogeneous sample and active learning/heterogeneous sample

The same uninstructed differential conditioning paradigm was used in both the Active Learning/Homogeneous Sample (Kruse et al., 2018) and Active Learning/Heterogeneous Sample (Kruse et al., 2020). In each trial, the subject was presented with a CS+ or CS- (blue or yellow rectangle) and then with a target (white square), upon which they were instructed to press a button as quickly as possible. Reactions within target presentation time were rewarded with 50 cents (UCS) only if a CS+ was presented before the target (timing of the target was predetermined, so that approx. 62% of all CS+ trials were rewarded). Fast reactions after a CS- were never rewarded. Participants were instructed to pay attention to possible contingencies before the task and received the money they won after scanning. The paradigm included 21 CS+ and 21 CS- trials. The first two trials (always one CS+ and one CS-) were excluded from further analyses, since learning could not have taken place yet, leaving 20 CS+ and CS- trials each per subject. For more detailed information about the paradigm please see the original publications for the Active Learning/Homogeneous Sample (Kruse et al., 2018) and the Active Learning/Heterogeneous Sample (Kruse et al., 2020). See also Fig. 1 for graphical representation of the task.

2.2.2. Passive learning/heterogeneous sample

In the Passive Learning/Heterogeneous Sample (Tapia León et al., 2019), an instructed differential conditioning paradigm without any behavioral reaction component was used. Participants were presented with a CS+ or CS- (blue or yellow rectangles) followed by feedback about reward/no reward. Half of the CS+ trials were rewarded with 50 cents (UCS) while the CS- was never rewarded. Participants were instructed about the relationships between CS and UCS before the task and received the money from the experiment after leaving the scanner. The paradigm included 20 CS+ trials and 20 CS- trials. For more detailed information about the paradigm see the original publication for the Passive Learning/Heterogeneous Sample (Tapia León et al., 2019). See also the right half of Fig. 1 for graphical representation of the task.

2.3. Appetitive sample data

MRI images for all samples were acquired using the same 3 T whole-body tomograph (Siemens Prisma). Preprocessing and first level analyses were performed using Matlab and Statistical Parametric Mapping (SPM 12) implemented in Matlab R2012a (The MathWorks Inc.). Event-related general linear models in each sample included appCS+ and appCS- in addition to other task and nuisance regressors. All following analyses use appCS+, appCS- as well as appCS+ > appCS- first level contrast images from these models. For detailed information on data acquisition, preprocessing and first level analysis, please see the supplementary information (S1 and S2) or the original sample publications (Active Learning/Homogeneous Sample: Kruse et al. 2018; Active Learning/Heterogeneous Sample: Kruse et al. 2020; Passive Learning/Heterogeneous Sample: Tapia León et al. 2019).

For this study, we additionally created group level contrast maps using paired t-tests on CSF-scaled and winsorized appCS+ and appCS- maps with custom code available from the authors' website (https://canlab.github.io; CANlab, code used for this publication available from https://github.com/s-kline/aversive-appetitive-conditioning). These were only used for visualization purposes (see Fig. 2) and not part of any subsequent analysis.

Finally, to judge how well activation data can be distinguished between appCS+ and appCS- condition without the aversive pattern, we performed multivariate predictive modeling analyses on the appetitive data only using custom code (https://canlab.github.io; CANlab, 2020, https://github.com/s-kline/aversive-appetitive-conditioning). In each conditioning sample, a classifier was trained and tested to distinguish between appCS+ and appCS- using whole-brain Support Vector Machines (Burges, 1998; Gramfort et al., 2013). We used 5-fold cross-validation blocked by subject (i.e., leaving out all images from a particular participant together), which allows every subject to serve as both training and test data at one point. The classifiers were trained on whole-brain appCS+ > appCS- first level contrast images masked with a gray matter mask. Each SVM model resulted in a pattern of weights of each voxel predicting the appCS+ or appCS- stimulus presentation (appCS+ > appCS- predictive weight map) and an intercept (offset) value. Bootstrap resampling (with 5,000 bootstrap samples; see also Wager et al. 2013) was used to estimate voxel-wise p-values for each predictive weight map. We tested for significant clusters in the predictive weight maps thresholded at P = .05, FDR (false discovery rate)-corrected.

2.4. Aversive conditioning pattern

For the aversive conditioning pattern, we used a whole brain pattern which discriminates within aversive conditioning paradigms between CS+ (avCS+) and CS- (avCS-). This avCS+ > avCS- pattern was the result of a meta-analysis of 27 independent fear conditioning data sets (total subjects N = 677, 54% male; Fullana et al. 2016). Specifically, Fullana et al. computed functional activation differences between avCS+ and avCS- for each study, either from original contrast maps or

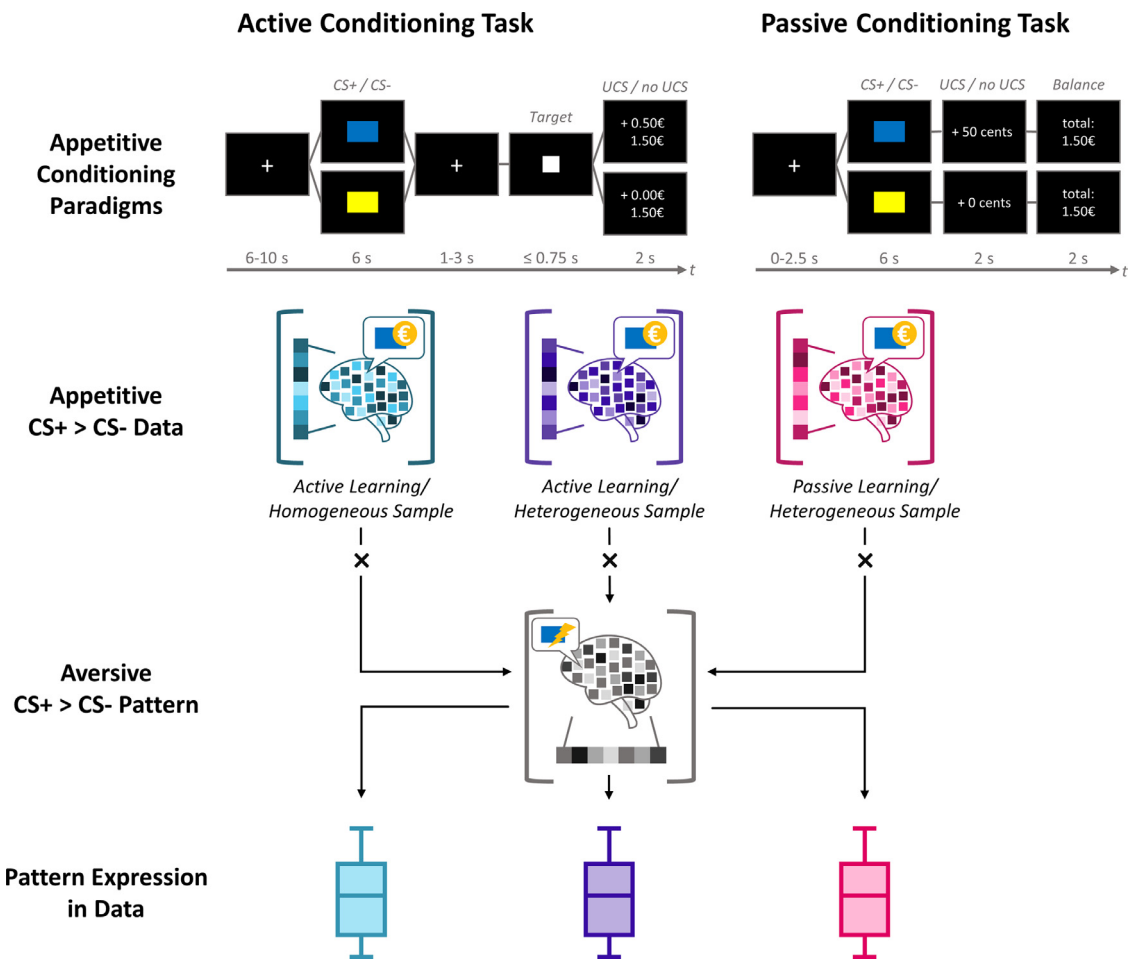

Fig. 1. Methods Summary

Note: Both active samples underwent the same active appetitive conditioning task. In each trial, subjects could win 50 cents with a fast reaction to the target only if a CS+ was shown before. The Passive Sample underwent a passive appetitive conditioning task. Subjects were shown CS+ and CS- and subsequent wins of 50 cents or nothing. Activation maps related to appetitive CS+ and CS- presentation averaged over the whole task were computed for each subject in each sample. The aversive conditioning pattern was applied to these subject-specific maps using cosine similarity metric. The pattern expression values reflect the magnitude of similarity between two normalized image vectors.

the peak coordinates reported in the studies. They then created a brain map of the effect size of the difference between the two conditions for each study using AES-SDM software (www.sdmproject.com/) and with these maps conducted a voxel-wise random-effects meta-analysis with weighting for sample size and variance. Fullana et al. (2016) found several large bilateral clusters demonstrating consistently significant functional activations during aversive conditioning (avCS+ > avCS-) including anterior insular cortex, NAcc, caudate nucleus, dACC and lateral cerebellum. Most of the included studies used electric shocks as UCS and simple geometric shapes as CS. The whole brain map of *z*-values associated with the difference between avCS+ and avCS- is available on Neurovault (https://identifiers.org/neurovault.collection:2472). We obtained this map of z-values from Neurovault and used it as the pattern associated with avCS+ > avCS- in our similarity analysis.

2.5. Similarity analysis

We followed the same analysis steps in each sample, using custom code available from the authors' website (https://canlab.github.io; CANlab; code used for this publication available from https://github.com/s-kline/aversive-appetitive-conditioning): (i) First, we computed pattern expression scores in the whole brain appCS+ > appCS- contrast images. (ii) Second, we computed the pattern expression in each of the ROIs NAcc, caudate nucleus, putamen amygdala, thalamus, insula and cerebellum. (iii) Finally, we computed pattern expression scores in the separate appCS+ and appCS- activation maps, which were then used in a classification analysis to test if we can distinguish appCS+ from appCS- condition based on these scores. The significance threshold for all tests was $P < .05$.

(i) To apply the pattern to our data, we initially resampled the pattern map to the space of the functional data using trilinear interpolation. Then, we used cosine similarity metric to assess the degree of similarity between the avCS+ > avCS- pattern and the individual unthresholded appCS+ > appCS- contrast image of each subject: For every subject of each of the three appetitive conditioning samples, we calculated a pattern expression score, which measures the similarity of the contrast image to the aversive conditioning pattern. As pattern expression score, we used the cosine similarity metric, which indicates to what extent the pattern image vector and the data image vector from one participant point in the same direction (Bisandu et al., 2019; Bobadilla-Suarez et al., 2020; van Oudenhove et al., 2020). For each appetitive sample participant, we calculated the dot product between the avCS+ > avCS- pattern image and their appCS+ > appCS- contrast image and divided it by the product of the two image vectors length, normalizing the result. Cosine similarity can range from 1 (indicating exact similarity, i.e. exactly the same direction of the vectors) over 0 (indicating no relation, orthogonal

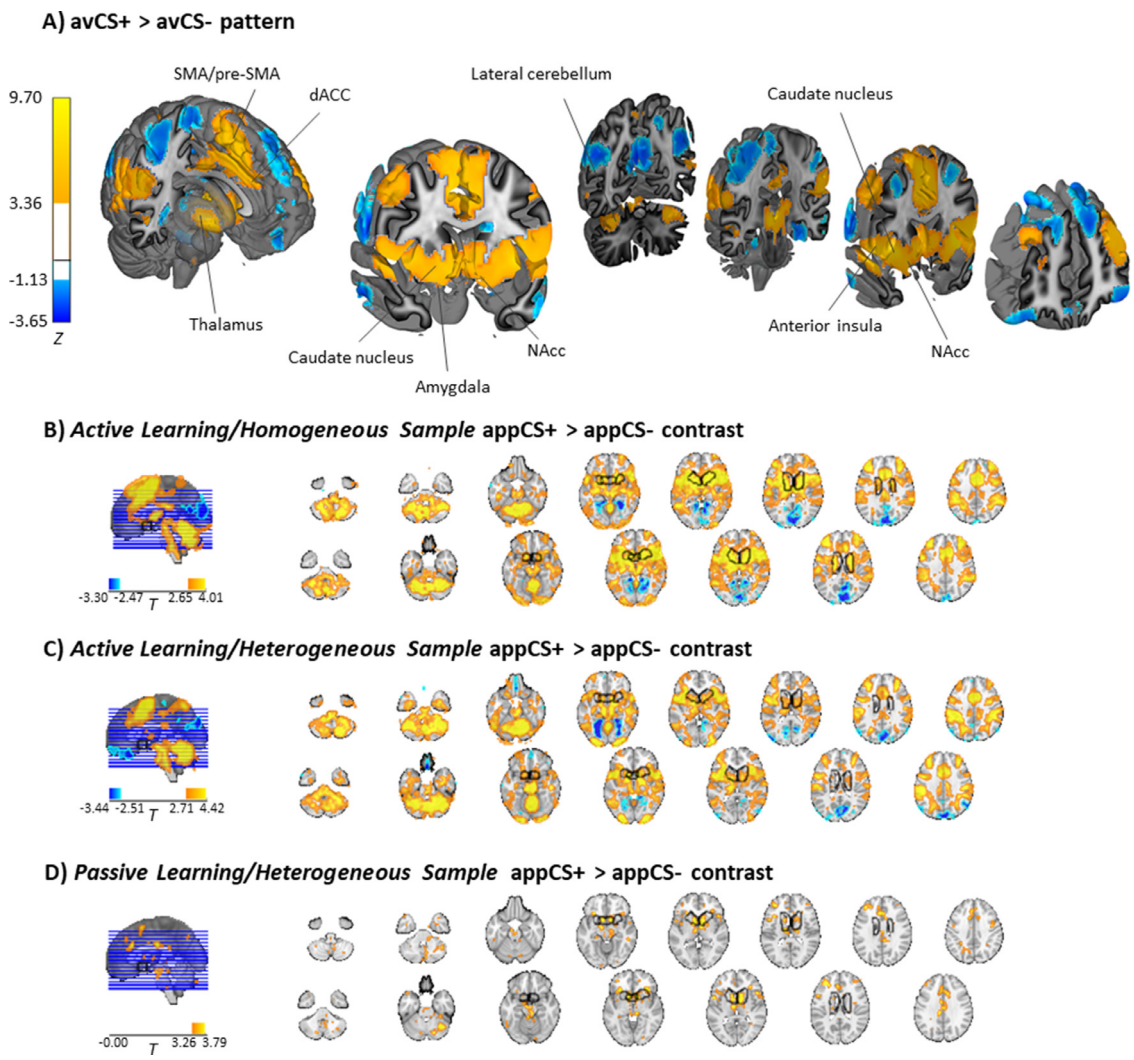

Fig. 2. Brain maps of aversive conditioning pattern and appetitive conditioning data
Note: Pattern related to aversive conditioning from meta-analysis (A). Weight map consisting of z-values is displayed on 4 coronal slices and two central cutaways showing the basal ganglia with region labels (SMA: supplementary motor area). The pattern is thresholded at $P < .005$, cluster size > 10, see Fullana et al. (2016) for details. Main effects of appetitive CS+ versus appetitive CS- in Samples (B, C, D). Contrast maps are the result of a paired t-test between CSF-scaled and winsorized activation maps of appCS+ and appCS- conditions, thresholded at $P < .05$ FDR-corrected. Midline sagittal and two rows of axial slices are shown for each sample, black outlines indicating NAcc and caudate nucleus. Anatomical images were adapted from the 7T high-resolution atlas of Keuken et al. (2014).

vectors) to -1 (indicating complete inversion, exactly opposite vector direction). Thus, in our analysis, positive cosine similarity (between 0 and 1) results when positive contrast values (appCS+ > appCS-) are found in voxels that are also positive in the aversive conditioning pattern. In accordance with that, positive cosine similarity also results when negative contrast values (appCS+ < appCS-) are found in voxels that are also negative in the aversive conditioning pattern. Equivalently, negative cosine similarity (between 0 and -1) results when positive contrast values are found in voxels that are negative in the aversive conditioning pattern and vice versa. Using this approach resulted in one pattern expression score per participant, which indicated the similarity between individual appetitive conditioning contrast images and the aversive conditioning pattern. Finally, we tested whether the appetitive conditioning contrast images were significantly similar to the aversive conditioning pattern using standard binomial tests with t-statistics, i.e. if cosine similarity was significantly different from 0.

(ii) For the ROI analysis, we masked the appCS+ > appCS- contrast images with anatomical masks for the NAcc (from the SPM anatomy toolbox), caudate nucleus, putamen (both from striatum parcellation by Pauli et al. 2016), amygdala (from the SPM anatomy toolbox), thalamus, insula (both from Harvard Oxford Atlas) and cerebellum (from Diedrichsen et al. 2009). This resulted in seven new images that only contained data in the voxels encompassed by the respective ROI. We then calculated the pattern expression scores in these images, which restricts the analysis to only the voxels within the ROI for both contrast image and pattern. Otherwise, we employed the same steps, cosine similarity metric and significance test as for the whole brain analysis described under (i).

(iii) We also computed cosine similarity of the avCS+ > avCS- pattern to the separate appCS+ and appCS- activation maps of each subject to use for classification analysis. This resulted in two pattern expression scores per participant, one indicating similarity of the pattern with appCS+ and the other one indicating similarity of the pattern with appCS-. To assess how the pattern expression scores for appCS+ and appCS- images differed from each other, we tested whether they could accurately predict the condition label appCS+ or appCS-. For this purpose, we computed forced-choice classification performance where the image with the higher avCS+ > avCS- pattern expression scores is labeled as appCS+ and the image with the smaller pattern expression score is labeled as appCS- using receiver operating characteristics (ROC; for

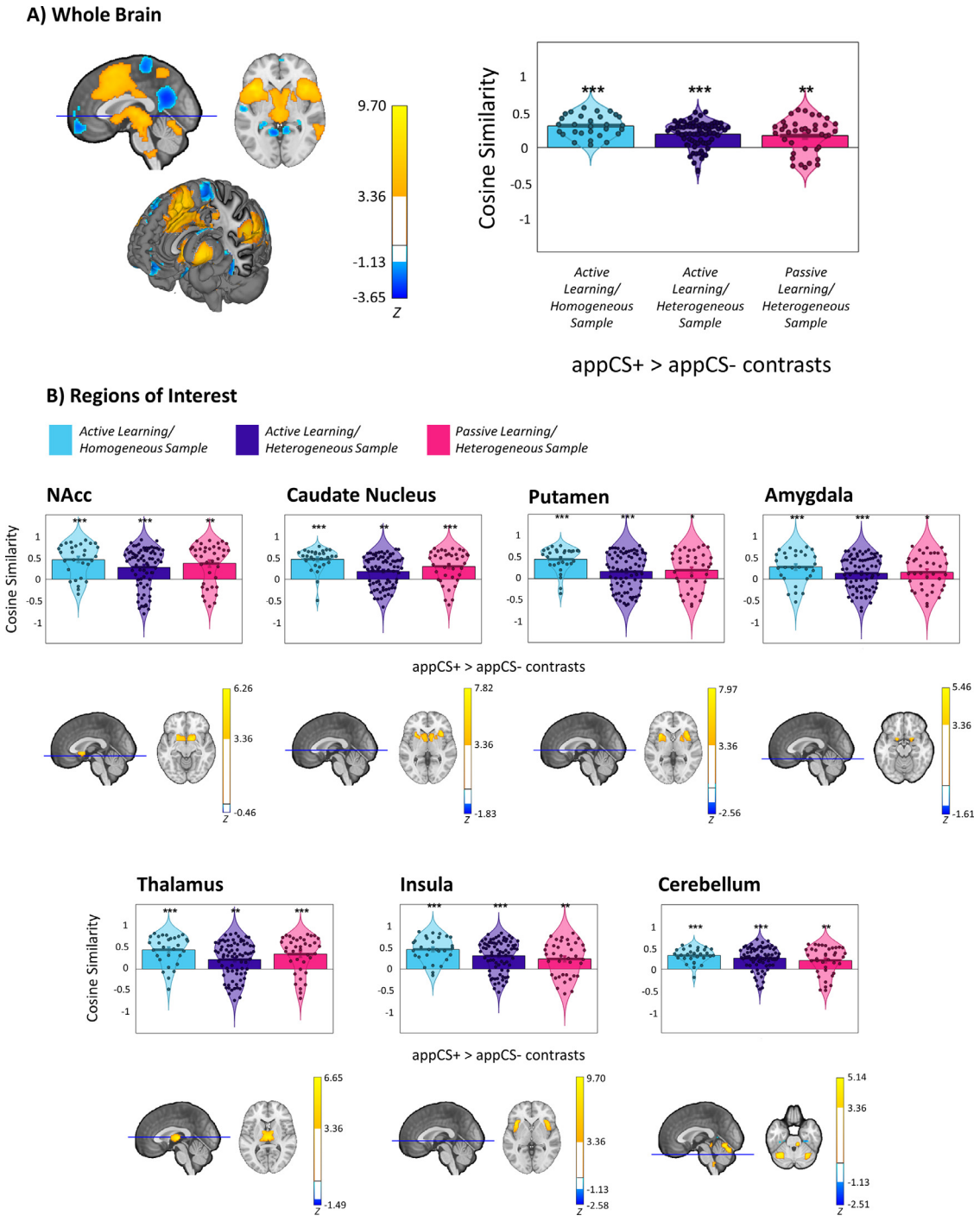

Fig. 3. Similarity between Aversive Conditioning Pattern and Appetitive Conditioning Data
Note: Results of similarity analysis for appetitive conditioning data in A) whole brain, and B) Regions of Interest NAcc, caudate nucleus, putamen, amygdala, thalamus, insula and cerebellum. For each region, the aversive conditioning pattern is shown mapped onto canonical anatomical sections (axial slice indicated by the line on mid-sagittal slice) and for the whole brain also onto respective brain cutaways adapted from the 7T high-resolution atlas of Keuken et al. (2014). The pattern is thresholded at $P < .005$, cluster size > 10, see Fullana et al. (2016) for details. Bar plots show the cosine similarity between aversive conditioning pattern and appetitive conditioning contrasts with each subject as a dot, violin plots illustrating the data distribution and error bars indicating standard error of means. *** indicates $P < .001$, ** indicates $P < .01$.

an introduction see Tharwat, 2021). We report accuracy measures and statistics of this classification based on the pattern expression values.

2.6. Control analyses

To support our assumption that similarity between aversive and appetitive conditioning is not solely driven by a common level of cognitive demand or emotional arousal features of both tasks, we performed control analyses. Specifically, pattern expression of other published whole brain multivariate patterns related to these concepts in the appetitive conditioning data were assessed. These were multivariate signatures related to cognitive control (Kragel et al., 2018), cognitive demand in a stroop task (Silvestrini et al., 2020), negative affect induced by pictures (Chang et al., 2015), as well as fearfulness and surprise induced

Table 2
Mean cosine similarity of avCS+ > avCS- pattern to appCS+ > appCS- contrast for whole brain and ROIs with standard error, statistics and effect size.

Region	Dataset	Cosine similarity	SE	T	p	Cohens d
Whole	Active Learning/Homogen.	0.304	0.028	10.85	<.001	2.02
Brain	Active Learning/Heterogen.	0.184	0.021	8.76	<.001	1.01
	Passive Learning/Heterogen.	0.160	0.039	4.13	<.001	0.67
NAcc	Active Learning/Homogen.	0.452	0.071	6.40	<.001	1.19
	Active Learning/Heterogen.	0.274	0.054	5.08	<.001	0.58
	Passive Learning/Heterogen.	0.360	0.077	4.68	<.001	0.76
Caudate	Active Learning/Homogen.	0.459	0.048	9.53	<.001	1.77
Nucleus	Active Learning/Heterogen.	0.179	0.041	4.39	<.001	0.50
	Passive Learning/Heterogen.	0.294	0.060	4.90	<.001	0.79
Putamen	Active Learning/Homogen.	0.444	0.052	8.52	<.001	1.58
	Active Learning/Heterogen.	0.159	0.049	3.24	.002	0.37
	Passive Learning/Heterogen.	0.191	0.072	2.66	.012	0.43
Amygdala	Active Learning/Homogen.	0.278	0.068	4.08	<.001	0.76
	Active Learning/Heterogen.	0.131	0.045	2.93	.005	0.34
	Passive Learning/Heterogen.	0.158	0.065	2.43	.020	0.40
Thalamus	Active Learning/Homogen.	0.443	0.064	6.93	<.001	1.29
	Active Learning/Heterogen.	0.212	0.048	4.42	<.001	0.51
	Passive Learning/Heterogen.	0.339	0.069	4.94	<.001	0.80
Insula	Active Learning/Homogen.	0.450	0.053	8.56	<.001	1.59
	Active Learning/Heterogen.	0.297	0.042	7.03	<.001	0.81
	Passive Learning/Heterogen.	0.229	0.064	3.55	.001	0.58
Cerebellum	Active Learning/Homogen.	0.309	0.031	9.82	<.001	1.82
	Active Learning/Heterogen.	0.247	0.030	8.28	<.001	0.95
	Passive Learning/Heterogen.	0.197	0.053	3.74	<.001	0.61

by music and films (Kragel and LaBar, 2015) available from the authors' website (https://canlab.github.io; CANlab). We computed expression of these patterns in each sample and tested for significance same as for the aversive pattern (see Section 2.5). If the similarity between aversive conditioning pattern and appetitive conditioning data is at least somewhat specific to conditioning, the similarity to these control patterns should be smaller in comparison. To test this, we performed paired t-tests to compare control pattern similarity and aversive conditioning pattern similarity with the appetitive conditioning data.

3. Results

3.1. Aversive pattern expression in appetitive contrast data

In line with our expectations, the aversive pattern was expressed significantly in the contrast images of every sample (all $p < .001$). Pattern expression was largest in the Active Learning/Homogeneous Sample with a mean cosine similarity of 0.304 (SE = 0.028, $t = 10.85$) and a very large effect size (Cohens $d = 2.02$). In the Active Learning/Heterogeneous Sample, pattern expression was moderate (cosine similarity = 0.184, SE = 0.021, $t = 8.76$, $d = 1.01$), but statistics and effect size of the similarity were still high; higher than in Passive Learning/Heterogeneous Sample (cosine similarity = 0.160, SE = 0.039, $t = 4.13$, $d = 0.67$).

As expected, pattern expression scores were also significantly large in all a priori ROIs. We found the highest scores in the striatal regions, thalamus and insula, moderately high scores in the cerebellum and moderate scores in the amygdala (for detailed statistics, see Table 2). Cosine Similarities between avCS+ > avCS+ pattern and the appCS+ > appCS- contrasts in the independent datasets are presented in Fig. 3 for whole brain and ROI data. For visual comparison, the aversive pattern as well as group contrast maps are shown in Fig. 2.

3.2. Classification of appCS+ versus appCS- by pattern expression

We computed pattern expression scores for the avCS+ > avCS- pattern in the separate appCS+ and appCS- conditions (see Fig. 4A and supplemental Table 1) to use for classification analysis. Classification results indicated that the aversive conditioning pattern could distinguish appCS+ from appCS- images accurately in every sample (classification performance in all three samples is presented in Fig. 4B). Forced choice classification effect size was largest in the Active Learning/Homogeneous Sample (100% accuracy, $d = 2.08$). The effect was also large in the Active Learning/Heterogeneous Sample (84% accuracy, $d = 1.05$) and moderate in the Passive Learning/Heterogeneous Sample (74% accuracy, $d = 0.80$). Importantly, the classification accuracies of the pattern for appCS+ versus appCS- were significantly above chance in all samples (all $p < .05$, see Table 3). These results are in line with the previous appCS+ > appCS- pattern expression results. Mean pattern expression scores were high in the appCS+ condition, supporting the notion that appCS+ activation data and avCS+ > avCS- pattern are highly similar. The pattern was also significantly expressed in the appCS- condition in every sample, likely due to basic similarities of CS+ and CS- conditions in both aversive and appetitive conditioning.

3.3. Control pattern expression in appetitive contrast data

As expected, all control patterns showed lower pattern expression in the appetitive sample data than the aversive conditioning pattern with all mean cosine similarity values < 0.08 (see Table 4 for detailed results). Only the pattern related to cognitive demand in a stroop task (Silvestrini et al., 2020) was significantly expressed in the appetitive sample data. This is probably because the stroop task has basic visual features and reaction demands in common with the appetitive conditioning paradigms. The pattern related to fearfulness (Kragel and LaBar, 2015) was significantly negatively expressed in the Active Learning/Homogenous Sample. In line with our expectations, the aversive conditioning pattern was expressed more strongly in appetitive conditioning data than any pattern related to cognitive demands and emotion processing (all p<.05 in paired t-tests, see Table 4).

3.4. Appetitive conditioning SVM classification

For all conditioning samples, we obtained predictive weight maps through support vector machine (SVM) classification (shown in supplemental Fig. 1A–C). The classifier trained on the Active Learning/Homogeneous Sample performed with 100% accuracy and a large effect size ($d = 2.62$), indicating that the cross-validated SVM scores

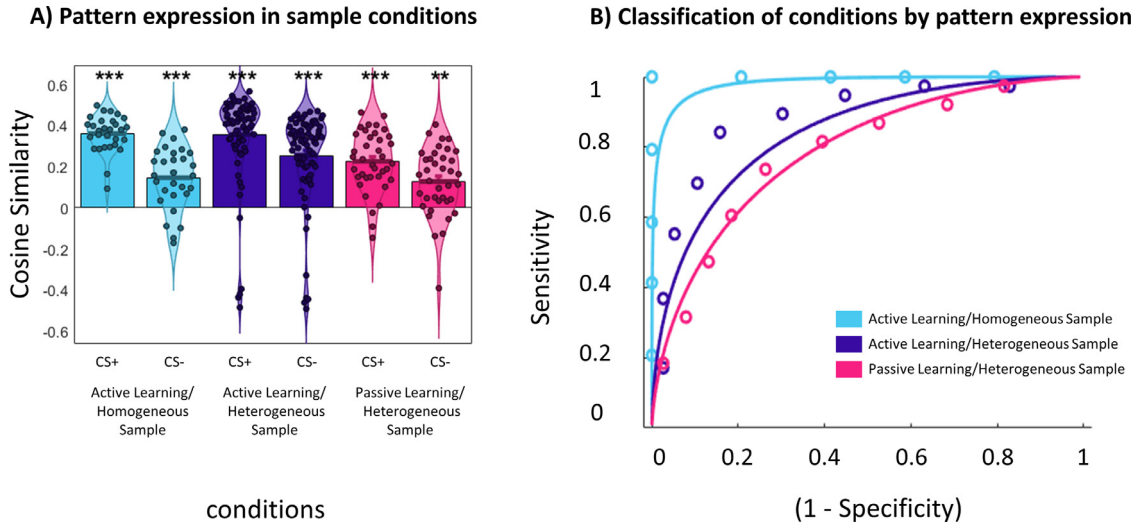

Fig. 4. Classification of Appetitive Data by Aversive Pattern
Note: (A) Bar plot showing cosine similarity between aversive pattern and appetitive sample conditions with each subject as a dot, violin plots illustrating the data distribution and error bars indicating standard error of means. (B) ROC plot showing aversive pattern performance on appCS+ vs. appCS- classification of data from all three samples. The threshold for classification, calculated with optimal balanced error rate was 0.0 for all samples. *** indicates $P < .001$, ** indicates $P < .01$.

Table 3
Performance of avCS+ > avCS- pattern classifying appCS+ versus appCS- conditions in three datasets. Accuracy with standard error (SE), specificity and sensitivity with confidence interval (CI) are presented to demonstrate the performance of the patterns using forced choice classification. Effect size indicates Cohen's d. *** indicates $p < .001$, ** indicates $p < .01$

Dataset	Accuracy (%)		Specificity (%)		Sensitivity (%)		Effect Size
		SE		CI		CI	
Active Learning/Homogen.	100***	0.0	100	100-100	100	100-100	2.08
Active Learning/Heterogen.	84***	4.0	84	76-91	84	76-91	1.05
Passive Learning/Heterogen.	74**	7.1	74	58-86	74	59-88	0.80

were higher for appCS+ than appCS- in every subject. The classifier trained on the Active Learning/Heterogeneous Sample performed moderately accurate (accuracy = 91%, d = 1.93) as did the classifier trained on the Passive Learning/Heterogeneous Sample (accuracy = 89%, $d = 1.58$). Accuracy was significantly above chance level (50%) as assessed with a binomial test for all classifiers ($P < .001$). Specificity, sensitivity, effect size, and accuracy for all three samples are presented in supplemental Table 1 (see also Supplemental Fig. 1D).

In the Active Learning/Homogeneous Sample predictive weight map, clusters significantly predicting the appCS+ versus appCS- condition were found. Clusters with positive effects (i.e. associated with the appCS+ compared to appCS-) were located in the NAcc, caudate nucleus, putamen, brainstem, cerebellum and somatomotor cortex. The weight maps of the Active Learning/Heterogeneous Sample and the Passive Learning/Heterogeneous Sample were predictive over the whole brain. There were no clusters limited to specific brain regions, which reached significance (all $P > .05$, FDR-corrected). All significant clusters from Active Learning/Homogeneous Sample are shown in supplemental Table 2.

4. Discussion

The goal of this study was to integrate neuroimaging findings from aversive with appetitive conditioning paradigms to empirically identify commonalities, and to show the feasibility of cross-paradigm integration with this example. Similarity of these processes in the brain has already been hypothesized but based mainly on qualitative literature reviews (Menon and Uddin, 2010; Moscarello and LeDoux, 2013; Seeley et al., 2007; Stefanova et al., 2020). We wanted not only to quantitatively assess the aversive pattern expression in an appetitive sample but also to determine if results would generalize across multiple appetitive conditioning datasets with differences in task, procedure, instruction, and sample makeup. To address this question, we analyzed three independent previously published appetitive conditioning datasets: The Active Learning/Homogeneous Sample (Kruse et al., 2018), the Active Learning/Heterogeneous Sample (Kruse et al., 2020) and the Passive Learning/Heterogeneous Sample (Tapia León et al., 2019). The aversive conditioning pattern was expressed significantly in the activation maps of all three appetitive conditioning datasets. Furthermore, we were able to accurately classify appCS+ from appCS- in all samples using the aversive pattern. These results provide robust empirical evidence for aversive and appetitive learning processes sharing common neural mechanisms.

The results are in line with previous research (Carter et al., 2009; Lake et al., 2019; Sankar et al., 2019) and we are now able to quantify the long-assumed similarity of aversive and appetitive learning processes at a neural level (Menon and Uddin, 2010; Moscarello and LeDoux, 2013; Seeley et al., 2007; Stefanova et al., 2020). Our results suggests that the activation difference between avCS+ and avCS- contains neural activation which is independent of UCS valence. This common activation might represent the acquired salience of both CS+ (Ogawa et al., 2013; Treviño, 2015). Furthermore, our results are consistent with a regional overlap in activation related to both negative and positive affective processing (Satpute et al., 2015) and appetitive and aversive prediction errors (Corlett et al., 2022). Animal studies recording activity from single cells and neuron populations have also shown that while appetitive and aversive CS+ may evoke distinct neural responses, they are often co-localized in the same anatomical areas (O'Neill et al., 2018; Shabel and Janak, 2009; Tye et al., 2010; Xiu et al., 2014). Thus, the commonalities we found may also reflect valence-specific activation in the same voxels. Finally, no pattern re-

Table 4
Mean cosine similarity of control patterns to appCS+ > appCS- contrast for whole brain with standard error, statistics and effect size (columns 1-7). Comparison between similarity of the control pattern and similarity of the avCS+>avCS- pattern with the respective appCS+ > appCS- contrast is shown with statistics (columns 8-9).

							Comparison with mean cosine similarity of avCS+>avCS- pattern	
Pattern	Dataset	Cosine similarity	SE	T	p	Cohens d	T	p
Cognitive Control, Kragel, Kano et al. (2018)	Active Learning/Homogen.	-0.009	0.010	-0.91	0.371	-0.17	9.18	<.001
	Active Learning/Heterogen.	-0.002	0.005	-0.31	0.758	-0.04	8.38	<.001
	Passive Learning/Heterogen.	-0.007	0.009	-0.76	0.452	-0.12	3.99	<.001
Cognitive Demand (Stroop; Silvestrini et al. 2020)	Active Learning/Homogen.	0.076	0.011	7.03	<.001	1.31	8.87	<.001
	Active Learning/Heterogen.	0.050	0.007	7.07	<.001	0.81	7.53	<.001
	Passive Learning/Heterogen.	0.054	0.012	4.60	<.001	0.75	3.56	.001
Fearful, Kragel and LaBar (2015)	Active Learning/Homogen.	-0.029	0.008	-3.70	<.001	-0.69	12.26	<.001
	Active Learning/Heterogen.	-0.008	0.005	-1.63	.108	-0.19	8.81	<.001
	Passive Learning/Heterogen.	0.008	0.007	1.07	.291	0.17	3.87	<.001
Surprise, Kragel and LaBar (2015)	Active Learning/Homogen.	<0.000	0.010	-0.02	.988	-0.00	11.32	<.001
	Active Learning/Heterogen.	<0.000	0.007	-0.04	.968	-0.01	8.49	<.001
	Passive Learning/Heterogen.	0.010	0.009	1.13	.264	-0.18	3.77	<.001
Picture Induced Negative Affect, Chang et al. (2015)	Active Learning/Homogen.	0.002	0.004	0.56	.583	0.10	10.67	<.001
	Active Learning/Heterogen.	-0.001	0.003	-0.25	.805	-0.03	8.68	<.001
	Passive Learning/Heterogen.	0.005	0.004	1.33	.192	0.22	3.99	<.001

lated to cognitive task demands or affective processing was expressed as highly in the appetitive conditioning data as the aversive conditioning pattern. This indicates that their similarity may be in part specific to the underlying motivational learning processes and not exclusively due to common task demands or basic sensory features.

In addition to similarity over the whole brain, we also found high similarity in the NAcc, caudate nucleus, putamen amygdala, thalamus, insula and cerebellum ROIs. This fits well with the roles of NAcc in reward and loss anticipation (Oldham et al., 2018), caudate nucleus in processing motivational values of actions (Balleine and O'Doherty, 2010), putamen in stimulus-response associations (Everitt and Robbins, 2013), amygdala in representing the CS-UCS relationship (Moscarello and LeDoux, 2013) and the cerebellum in predictive coding and motor responses (Lange et al., 2015) found in past conditioning studies. The thalamus is likely important as a sensory input region for the amygdala in both appetitive and aversive conditioning (Gründemann, 2021; Tye et al., 2008) while the insula may be involved in learning under uncertainty (Gorka et al., 2016; Morriss et al., 2019). Co-localization of aversive and appetitive learning responses in amygdala (O'Neill et al., 2018; Shabel and Janak, 2009; Tye et al., 2010) and striatal regions (Xiu et al., 2014) have already been found in animal studies and more recently, in a human fMRI meta-analysis (Corlett et al., 2022). Here, similarity was most notably high in NAcc and caudate nucleus, indicating that these striatal regions especially may be crucial for motivational salience learning. Further underpinning this interpretation, the SVM classifier trained on appetitive data only (Active Learning/Homogeneous Sample) also revealed clusters predicting appCS+ vs. appCS- in the NAcc, caudate nucleus and cerebellum (see supplemental Table 1). Importantly, while co-localized fMRI activation in these regions points to them being involved in appetitive as well as aversive learning, it may not necessarily indicate them performing the same functions during aversive and appetitive conditioning. For example, animal evidence suggests that activation in the NAcc shell indicates the motivational valence of both an appCS+ and an avCS+ arranged along a rostrocaudal gradient with more anterior activation indicating positive valence (approach signal) and more posterior activation indicating negative valence (avoidance signal; Berridge and Kringelbach 2015). Activation in the NAcc core most likely indicates an unsigned motivational salience signal based on the input it receives from the ventral tegmental areal (Bromberg-Martin et al., 2010). Thus, combined signals from the NAcc may be important for approach behavior in appetitive conditioning and avoidance behavior in aversive conditioning (Gentry et al., 2019) but signal motivational salience of the CS+ in both.

Our findings are particularly relevant since altered aversive and appetitive conditioning are considered the basis for psychological disorders characterized by excessive avoidance and approach behavior, respectively (Duits et al., 2015; Martin-Soelch et al., 2007). As of yet, very little is known about commonalities and overlaps between these disorder categories. Here, we have provided proof of concept for an approach which facilitates finding commonalities in such separate concepts. Further integration of data across more different affective learning paradigms and RDoC domains – and across patient samples - may help fill these knowledge gaps and pave the way towards transdiagnostic biomarkers (Insel, 2014; Woo et al., 2017).

Our results support the practicability of quantitative cross-paradigm integration. We found high whole brain similarity between aversive and appetitive CS+ > CS- contrasts in all samples (see Fig. 3). Effects were larger in some samples than others but present and significantly strong in all of them. These results demonstrate how empirical knowledge can be gained from disparate paradigms by quantifying their similarity. Using an existing software toolbox (https://canlab.github.io; CANlab), and an openly available meta-analysis (Fullana et al., 2016), we could effi-

ciently integrate our current appetitive data with a multitude of past aversive conditioning studies. Empirical cross-paradigm integration has rarely been done up until now – in this study, we could illustrate the feasibility of our approach. Considering the exponential increase in fMRI publications over the last two decades and the difficulties to collect large datasets at single institutions, data integration across studies is becoming an increasingly essential analysis tool. Tools such as these are much needed if we want to better understand the connections between the diverse published evidence and our own data. Here, using this method, we found remarkably high neural similarity with the aversive activation pattern in every appetitive sample included. This enables us to make conclusions not only about the neural similarity of aversive and appetitive learning itself but also about the generalizability of this similarity.

To verify and examine generalizability of the results, we included different appetitive conditioning samples. The otherwise often troublesome fact that many experiments on appetitive conditioning vary in details can be used to our advantage here. By including diverse studies, we can quantify the variance between them and thus try to evaluate how much those details actually affect results while at the same time assessing the generalizability of cross-paradigm similarity. In our analysis, we included three different samples, to examine how generalizable the integration results are. Results were significant across all three samples despite some small differences in effect sizes and classification accuracies. This variance in results may have been due to several reasons: (1) Smaller sample size and increased homogeneity may improve the estimation of experimental variance because of decreased noise. Some studies suggest that increased neural activation variance in conditioning paradigms can be due to hormone fluctuation differences in subjects assigned female at birth, depending on whether they use hormonal birth control (Merz et al., 2018). Thus, samples including mostly cis men may show especially low inter-subject variance. (2) A similar point can be made concerning a more standardized and strictly controlled study protocol – this likely reduces error variance. (3) Less instruction and increased action demands in an appetitive task may lead to it being more arousing overall and thus closer to the (presumably higher) arousal level in an aversive task. Both points (1) and (2) were given in Active Learning/Homogeneous Sample and (3) was a notable difference between Passive Learning and Active Learning samples. The influence of active versus passive task design on the similarity remains to be examined more closely but recent findings suggest that the common neurocircuitry between these types of tasks mirrors the commonalities we found here (mainly ventral and dorsal striatum; Corlett et al. 2022). SVM classification based on the appetitive data only also worked best in Active Learning/Homogeneous Sample, further illustrating how reduced inter-subject variance may improve modeling results generally. Thus, our results highlight the brain activation differences between appetitive conditioning experiments which vary only slightly in task and sample characteristics. At the same time, by integrating over a diversity of methods and samples, we could show that the similarities between patterns of activation associated with aversive and appetitive CS+ can be generalized across this diversity.

In all three samples, we also found the aversive pattern positively expressed in the appCS- condition to varying degrees (see Fig. 4A). Possible explanations for this include: First, basic similarity of the conditions – aversive pattern as well as both appCS+ and appCS- data likely contain activation related to general visual processing, attention etc. leading to a small baseline of similarity. Second, the appCS- may have acquired aversive properties since it signaled absence of a reward (Matsumoto and Hikosaka, 2009; Mollick et al., 2021). This is supported by a post-conditioning drop in appCS- valence ratings in the two Active Learning samples (Kruse et al., 2018, 2020). Part of the appCS- activation data may then reflect these aversive properties. However, the appCS+ condition was still more similar to the avCS+>avCS- pattern, indicating that the pattern primarily codes acquired salience rather than valence.

4.1. Limitations and future directions

Human fMRI data has limited spatial resolution compared to animal studies utilizing methods like single-unit recording or optical imaging. Thus, while we found the BOLD responses to aversive and appetitive CS quite similar at a voxel level, neuronal responses could still be dissociable at a much more precise spatial scale than possible to measure here (e.g. neuronal populations). Another limitation was that the appetitive samples differed in key details but were all collected at the same site. This may have made the overall procedures similar; the scanner itself, other facilities and some of the data collection staff were identical for all samples. Furthermore, while our appetitive learning paradigms are intentionally held similar to classical fear conditioning, the majority of appetitive learning paradigms used in human fMRI are more diverse than this (e.g. reinforcement learning with varying probabilities, risk-taking; Averbeck and Costa 2017, Sherman et al. 2018). The diversity of paradigms out there is a considerable resource that presents countless possibilities for further study with our integration approach. Aversive conditioning data could be integrated with a broader range of appetitive learning datasets that have more procedural and task variance between them. Doing this will enable us to more closely narrow down the factors involved in their similarity. To better understand dissimilarities, integration could also be done with increasingly different paradigms, for example starting with affectively neutral associative conditioning. This could also answer the open question, whether the similarities found here are due to both paradigms involving learning contexts or if the emotional context that they share is more important. Another open question is how the similarity between appetitive and aversive conditioning is mediated by using primary versus secondary UCS. Future studies could disentangle this effect from affective valence by repeating the analysis with more primary appetitive UCS such as food or water instead of money as a secondary reward. Further integration with more different paradigms may reveal how much of the similarity found here can be attributed to common basic features of most fMRI tasks, such as sensory processing, attention and motor action. We included control patterns related to basic cognitive and emotional processing as a first step in this validation process. An alternative to using an aversive meta-analysis pattern as we did here would be to train an aversive conditioning classifier and testing it in appetitive data. Having a sample where each participant performs an appetitive as well as aversive conditioning paradigm would also enable to train aversive conditioning patterns individually for each participant and testing similarity with appetitive conditioning data in the same individual. Such finer-grained aversive patterns may be more precise predictors and also illuminate possible individual differences concerning the similarity between aversive and appetitive learning. Finally, since there are different avenues of integrating data (e.g. principal component analysis), future work could also expand the methods some more to see if more information can be gained from other similarity metrics. Altogether, continuing rigorous cross-paradigm-integration may provide important clinical insights, as it allows to build on existing transdiagnostic approaches to mental health: For instance, the RDoC initiative seeks to characterize mental disorders by impaired functioning in various domains (such as fear and reward learning) rather than existing disorder categories (Insel, 2014). In this framework, fear and reward learning are separate domains, but cross-paradigm-integration findings could demonstrate the benefits of working not only within those domains but across them as well.

5. Conclusion

In conclusion, this study demonstrates the similarity of aversive and appetitive conditioning at fMRI pattern level across multiple independent appetitive datasets. These commonalities may have important implications for etiological models of fear- and reward-related disorders. Enabled by the open science movement and multivariate analysis methods, we could quantitatively integrate past evidence from one paradigm

with current data from another. Using the example of aversive and appetitive conditioning, we have demonstrated that this approach is not only viable but extremely valuable when trying to connect data from different paradigms. It presents an opportunity to integrate rather than compare past findings with current studies and thus make better use of the ever-growing body of fMRI studies.

Code and data availability

Data were analyzed using CANlab neuroimaging analysis tools available at https://github.com/canlab/, analysis code specific for this publication available from https://github.com/s-kline/aversive-appetitive-conditioning.

The data from the appetitive conditioning samples are available upon request from the corresponding author. The data are not publicly available due to ethical or privacy restrictions. The meta-analysis pattern is openly available at Neurovault (https://identifiers.org/neurovault.collection:2472).

Declaration of Competing Interest

None.

Credit authorship contribution statement

Sanja Klein: Writing – original draft, Conceptualization, Methodology, Formal analysis, Investigation. **Onno Kruse:** Writing – review & editing, Investigation. **Isabell Tapia León:** Writing – review & editing, Investigation. **Lukas Van Oudenhove:** Writing – review & editing, Methodology. **Sophie R. van 't Hof:** Writing – review & editing. **Tim Klucken:** Writing – review & editing, Resources, Funding acquisition. **Tor D. Wager:** Writing – review & editing, Resources, Software, Methodology. **Rudolf Stark:** Writing – review & editing, Resources, Funding acquisition.

Data Availability

Data will be made available on request.

Funding

The work of SK, TK, and RS on this article was carried out in the context of the Research Unit ACSID, FOR2974, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 411232260. The data collection for the Active Learning/Heterogeneous Sample (DFG KL2500/5-1) and for the Passive Learning/Heterogeneous Sample (DFG KL2500/4-1) was also funded by the German research association.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119594.

References

Averbeck, B.B., Costa, V.D., 2017. Motivational neural circuits underlying reinforcement learning. Nat. Neurosci. 20 (4), 505–512. doi:10.1038/nn.4506.

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. doi:10.1038/npp.2009.131, Official Publication of the American College of Neuropsychopharmacology.

Berridge, K.C., Kringelbach, M.L., 2015. Pleasure systems in the brain. Neuron 86 (3), 646–664. doi:10.1016/j.neuron.2015.02.018.

Bisandu, D.B., Prasad, R., Liman, M.M., 2019. Data clustering using efficient similarity measures. J. Stat. Manag. Syst. 22 (5), 901–922. doi:10.1080/09720510.2019.1565443.

Bobadilla-Suarez, S., Ahlheim, C., Mehrotra, A., Panos, A., Love, B., 2020. Measures of neural similarity. Comput. Brain Behav. 3 (4), 369–383. doi:10.1007/s42113-019-00068-5.

Breiter, H.C., Aharon, I., Kahneman, D., Dale, A., Shizgal, P., 2001. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron 30 (2), 619–639. doi:10.1016/S0896-6273(01)00303-8.

Bromberg-Martin, E.S., Matsumoto, M., Hikosaka, O., 2010. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron 68 (5), 815–834. doi:10.1016/j.neuron.2010.11.022.

Brooks, A.M., Berns, G.S., 2013. Aversive stimuli and loss in the mesocorticolimbic dopamine system. Trends Cogn. Sci. 17 (6), 281–286. doi:10.1016/j.tics.2013.04.001.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2, 121–167. https://canlab.github.io.

CANlab. CANlab neuroimaging analysis tools.

Carter, R.M., MacInnes, J.J., Huettel, S.A., Adcock, R.A., 2009. Activation in the VTA and nucleus accumbens increases in anticipation of both gains and losses. Front. Behav. Neurosci. 3. doi:10.3389/neuro.08.021.2009.

Chang, L.J., Gianaros, P.J., Manuck, S.B., Krishnan, A., Wager, T.D., 2015. A sensitive and specific neural signature for picture-induced negative affect. PLoS Biol. 13 (6), e1002180. doi:10.1371/journal.pbio.1002180.

Chase, H.W., Kumar, P., Eickhoff, S., Dombrovski, A.Y., 2015. Reinforcement learning models and their neural correlates: an activation likelihood estimation meta-analysis. Cogn. Affect. Behav. Neurosci. 15 (2), 435–459. doi:10.3758/s13415-015-0338-7.

Corlett, P.R., Mollick, J.A., Kober, H., 2022. Meta-analysis of human prediction error for incentives, perception, cognition, and action. Neuropsychopharmacology doi:10.1038/s41386-021-01264-3, Official Publication of the American College of Neuropsychopharmacology. Advance online publication.

Destoop, M., Morrens, M., Coppens, V., Dom, G., 2019. Addiction, anhedonia, and comorbid mood disorder. A narrative review. Front. Psychiatry 10 (311). doi:10.3389/fpsyt.2019.00311.

Diedrichsen, J., Balsters, J.H., Flavell, J., Cussans, E., Ramnani, N., 2009. A probabilistic MR atlas of the human cerebellum. Neuroimage 46 (1), 39–46. doi:10.1016/j.neuroimage.2009.01.045.

Duits, P., Cath, D.C., Lissek, S., Hox, J.J., Hamm, A.O., Engelhard, I.M., van den Hout, M.A., Baas, J.M.P., 2015. Updated meta-analysis of classical fear conditioning in the anxiety disorders. Depress. Anxiety 32 (4), 239–253. doi:10.1002/DA.22353.

Ernst, T.M., Brol, A.E., Gratz, M., Ritter, C., Bingel, U., Schlamann, M., Maderwald, S., Quick, H.H., Merz, C.J., Timmann, D., 2019. The cerebellum is involved in processing of predictions and prediction errors in a fear conditioning paradigm. Elife 8. doi:10.7554/eLife.46831.

Etkin, A., Wager, T.D., 2007. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. Am. J. Psychiatry 164 (10), 1476–1488. doi:10.1176/appi.ajp.2007.07030504.

Everitt, B.J., Robbins, T.W., 2013. From the ventral to the dorsal striatum: devolving views of their roles in drug addiction. Neurosci. Biobehav. Rev. 37 (9), 1946–1954. doi:10.1016/j.neubiorev.2013.02.010, Pt A.

Fullana, M.A., Harrison, B.J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., Radua, J., 2016. Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. Mol. Psychiatry 21 (4), 500–508. doi:10.1038/mp.2015.88.

Gentry, R.N., Schuweiler, D.R., Roesch, M.R., 2019. Dopamine signals related to appetitive and aversive events in paradigms that manipulate reward and avoidability. Brain Res. 1713, 80–90. doi:10.1016/j.brainres.2018.10.008.

Gorka, S.M., Nelson, B.D., Phan, K.L., Shankman, S.A., 2016. Intolerance of uncertainty and insula activation during uncertain reward. Cogn. Affect. Behav. Neurosci. 16 (5), 929–939. doi:10.3758/s13415-016-0443-2.

Gramfort, A., Thirion, B., Varoquaux, G., 2013. Identifying predictive regions from fMRI with TV-L1 prior. In: Davatzikos, C. (Ed.), Proceedings of the International Workshop on Pattern Recognition in Neuroimaging (PRNI). IEEE, Philadelphia, PA, USA, pp. 17–20. doi:10.1109/PRNI.2013.14 22-24 June 2013.

Gründemann, J., 2021. Distributed coding in auditory thalamus and basolateral amygdala upon associative fear learning. Curr. Opin. Neurobiol. 67, 183–189. doi:10.1016/j.conb.2020.11.014.

Heffley, W., Hull, C., 2019. Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. Elife (8) doi:10.7554/eLife.46764.

Herry, C., Johansen, J.P., 2014. Encoding of fear learning and memory in distributed neuronal circuits. Nat. Neurosci. 17 (12), 1644–1654. doi:10.1038/nn.3869.

Insel, T.R., 2014. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. Am. J. Psychiatry 171 (4), 395–397. doi:10.1176/appi.ajp.2014.14020138.

Kragel, P.A., Kano, M., van Oudenhove, L., Ly, H.G., Dupont, P., Rubio, A., Delon-Martin, C., Bonaz, B.L., Manuck, S.B., Gianaros, P.J., Ceko, M., Reynolds Losin, E.A., Woo, C.W., Nichols, T.E., Wager, T.D., 2018. Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. Nat. Neurosci. 21 (2), 283–289. doi:10.1038/s41593-017-0051-7.

Kragel, P.A., Koban, L., Barrett, L.F., Wager, T.D., 2018. Representation, pattern information, and brain signatures: From neurons to neuroimaging. Neuron 99 (2), 257–273. doi:10.1016/j.neuron.2018.06.009.

Kragel, P.A., LaBar, K.S., 2014. Advancing emotion theory with multivariate pattern classification. Emot. Rev. 6 (2), 160–174. doi:10.1177/1754073913512519.

Kragel, P.A., LaBar, K.S., 2015. Multivariate neural biomarkers of emotional states are categorically distinct. Soc. Cogn. Affect. Neurosci. 10 (11), 1437–1448. doi:10.1093/scan/nsv032.

Kruse, O., Klein, S., Tapia León, I., Stark, R., Klucken, T., 2020. Amygdala and nucleus accumbens involvement in appetitive extinction. Hum. Brain Mapp. 41 (7), 1833–1841. doi:10.1002/hbm.24915.

Kruse, O., Tapia León, I., Stalder, T., Stark, R., Klucken, T., 2018. Altered reward learning and hippocampal connectivity following psychosocial stress. Neuroimage 171, 15–25. doi:10.1016/j.neuroimage.2017.12.076.

Kruse, O., Tapia León, I., Stark, R., Klucken, T., 2017. Neural correlates of appetitive extinction in humans. Soc. Cogn. Affect. Neurosci. 12 (1), 106–115. doi:10.1093/scan/nsw157.

Lake, J.I., Spielberg, J.M., Infantolino, Z.P., Crocker, L.D., Yee, C.M., Heller, W., Miller, G.A., 2019. Reward anticipation and punishment anticipation are instantiated in the brain via opponent mechanisms. Psychophysiology 56 (8), e13381. doi:10.1111/psyp.13381.

Lange, I., Kasanova, Z., Goossens, L., Leibold, N., Zeeuw, C.I., de van Amelsvoort, T., Schruers, K., 2015. The anatomy of fear learning in the cerebellum: A systematic meta-analysis. Neurosci. Biobehav. Rev. 59, 83–91. doi:10.1016/j.neubiorev.2015.09.019.

Liverant, G.I., Sloan, D.M., Pizzagalli, D.A., Harte, C.B., Kamholz, B.W., Rosebrock, L.E., Cohen, A.L., Fava, M., Kaplan, G.B., 2014. Associations among smoking, anhedonia, and reward learning in depression. Behav. Ther. 45 (5), 651–663. doi:10.1016/j.beth.2014.02.004.

Mackintosh, N.J., 1975. A theory of attention: variations in the associability of stimuli with reinforcement. Psychol. Rev. 82 (4), 276–298. doi:10.1037/H0076778.

Martin-Soelch, C., Linthicum, J., Ernst, M., 2007. Appetitive conditioning: neural bases and implications for psychopathology. Neurosci. Biobehav. Rev. (3) 31. doi:10.1016/j.neubiorev.2006.11.002.

MATLAB R2012a, 2012. The MathWorks Inc.

Matsumoto, M., Hikosaka, O., 2009. Representation of negative motivational value in the primate lateral habenula. Nat. Neurosci. 12 (1), 77–84. doi:10.1038/nn.2233.

Mechias, M.L., Etkin, A., Kalisch, R., 2010. A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. Neuroimage 49 (2), 1760–1768. doi:10.1016/j.neuroimage.2009.09.040.

Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. Brain Struct. Funct. 214 (5-6), 655–667. doi:10.1007/s00429-010-0262-0.

Merz, C.J., Kinner, V.L., Wolf, O.T., 2018. Let's talk about sex … differences in human fear conditioning. Curr. Opin. Behav. Sci. 23, 7–12. doi:10.1016/j.cobeha.2018.01.021.

Mollick, J.A., Chang, L.J., Krishnan, A., Hazy, T.E., Krueger, K.A., Frank, G.K.W., Wager, T.D., O'Reilly, R.C, 2021. The neural correlates of cued reward omission. Front. Hum. Neurosci. 15, 615313. doi:10.3389/fnhum.2021.615313.

Morriss, J., Gell, M., van Reekum, C.M., 2019. The uncertain brain: a co-ordinate based meta-analysis of the neural signatures supporting uncertainty during different contexts. Neurosci. Biobehav. Rev. 96, 241–249. doi:10.1016/j.neubiorev.2018.12.013.

Moscarello, J.M., LeDoux, J.E., 2013. The contribution of the amygdala to aversive and appetitive pavlovian processes. Emot. Rev. 5 (3), 248–253. doi:10.1177/1754073913477508.

Ogawa, M., van der Meer, M.A.A., Esber, G.R., Cerri, D.H., Stalnaker, T.A., Schoenbaum, G., 2013. Risk-responsive orbitofrontal neurons track acquired salience. Neuron 77 (2), 251–258. doi:10.1016/j.neuron.2012.11.006.

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., Lorenzetti, V., 2018. The anticipation and outcome phases of reward and loss processing: a neuroimaging meta-analysis of the monetary incentive delay task. Hum. Brain Mapp. 39 (8), 3398–3418. doi:10.1002/hbm.24184.

O'Neill, P.K., Gore, F., Salzman, C.D., 2018. Basolateral amygdala circuitry in positive and negative valence. Curr. Opin. Neurobiol. 49, 175–183. doi:10.1016/j.conb.2018.02.012.

Pauli, W.M., O'Reilly, R.C., Yarkoni, T., Wager, T.D, 2016. Regional specialization within the human striatum for diverse psychological functions. Advance online publication. Proc. Natl. Acad. Sci. 113 (7), 1907–1912. doi:10.1073/PNAS.1507610113.

Popa, L.S., Ebner, T.J., 2018. Cerebellum, predictions and errors. Front. Cell. Neurosci. 12 (524). doi:10.3389/fncel.2018.00524.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., Nummenmaa, L., 2016. Discrete neural signatures of basic emotions. Cereb. Cortex 26 (6), 2563–2573. doi:10.1093/cercor/bhv086.

Sankar, A., Yttredahl, A.A., Fourcade, E.W., Mickey, B.J., Love, T.M., Langenecker, S.A., Hsu, D.T., 2019. Dissociable neural responses to monetary and social gain and loss in women with major depressive disorder. Front. Behav. Neurosci. 13 (149). doi:10.3389/fnbeh.2019.00149.

Satpute, A.B., Kang, J., Bickart, K.C., Yardley, H., Wager, T.D., Barrett, L.F., 2015. Involvement of sensory regions in affective experience: a meta-analysis. Front. Psychol. 6. doi:10.3389/fpsyg.2015.01860.

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., Greicius, M.D., 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. 27 (9), 2349–2356. doi:10.1523/JNEUROSCI.5587-06.2007.

Sehlmeyer, C., Schöning, S., Zwitserlood, P., Pfleiderer, B., Kircher, T., Arolt, V., Konrad, C., 2009. Human fear conditioning and extinction in neuroimaging: a systematic review. PLoS One 4 (6), e5865. doi:10.1371/journal.pone.0005865.

Shabel, S.J., Janak, P.H., 2009. Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. Proc. Natl. Acad. Sci. U.S.A. 106 (35), 15031–15036. doi:10.1073/pnas.0905580106.

Sherman, L., Steinberg, L., Chein, J., 2018. Connecting brain responsivity and real-world risk taking: Strengths and limitations of current methodological approaches. Dev. Cogn. Neurosci. 33, 27–41. doi:10.1016/j.dcn.2017.05.007.

Silvestrini, N., Chen, J.I., Piché, M., Roy, M., Vachon-Presseau, E., Woo, C.W., Wager, T.D., Rainville, P., 2020. Distinct fMRI patterns colocalized in the cingulate cortex underlie the after-effects of cognitive control on pain. Neuroimage 217, 116898. doi:10.1016/j.neuroimage.2020.116898.

Statistical Parametric Mapping (SPM 12), 2014. Wellcome Department of Cognitive Neurology, London, UK.

Stefanova, E., Dubljević, O., Herbert, C., Fairfield, B., Schroeter, M.L., Stern, E.R., Urben, S., Derntl, B., Wiebking, C., Brown, C., Drach-Zahavy, A., Loeffler, Kathrin, L, A., Albrecht, F., Palumbo, R., Boutros, S.W., Raber, J., Lowe, L, 2020. Anticipatory feelings: neural correlates and linguistic markers. Neurosci. Biobehav. Rev. 113, 308–324. doi:10.1016/j.neubiorev.2020.02.015.

Tapia León, I., Kruse, O., Stark, R., Klucken, T., 2019. Relationship of sensation seeking with the neural correlates of appetitive conditioning. Soc. Cogn. Affect. Neurosci. 14 (7), 769–775. doi:10.1093/scan/nsz046.

Tharwat, A., 2021. Classification assessment methods. Appl. Comput. Inform. 17 (1), 168–192. doi:10.1016/j.aci.2018.08.003.

Tovote, P., Fadok, J.P., Lüthi, A., 2015. Neuronal circuits for fear and anxiety. Nat. Rev. Neurosci. 16 (6), 317–331. doi:10.1038/nrn3945.

Treviño, M., 2015. Associative learning through acquired salience. Front. Behav. Neurosci. 9 (353). doi:10.3389/fnbeh.2015.00353.

Tye, K.M., Cone, J.J., Schairer, W.W., Janak, P.H., 2010. Amygdala neural encoding of the absence of reward during extinction. J. Neurosci. 30 (1), 116–125. doi:10.1523/JNEUROSCI.4240-09.2010.

Tye, K.M., Stuber, G.D., Ridder, B.de, Bonci, A., Janak, P.H., 2008. Rapid strengthening of thalamo-amygdala synapses mediates cue-reward learning. Nature 453 (7199), 1253–1257. doi:10.1038/nature06963.

van Oudenhove, L., Kragel, P.A., Dupont, P., Ly, H.G., Pazmany, E., Enzlin, P., Rubio, A., Delon-Martin, C., Bonaz, B., Aziz, Q., Tack, J., Fukudo, S., Kano, M., Wager, T.D., 2020. Common and distinct neural representations of aversive somatic and visceral stimulation in healthy individuals. Nat. Commun. 11 (1), 5939. doi:10.1038/s41467-020-19688-8.

van 't Hof, S.R., van Oudenhove, L., Janssen, E., Klein, S., Reddan, M.C., Kragel, P.A., Stark, R., Wager, T.D., 2021. The brain activation-based sexual image classifier (BASIC): a sensitive and specific fMRI activity pattern for sexual image processing. Cereb. Cortex doi:10.1093/cercor/bhab397, Advance online publication.

Wager, T.D., Atlas, L.Y., Lindquist, M.A., Roy, M., Woo, C.W., Kross, E., 2013. An fMRI-based neurologic signature of physical pain. N. Engl. J. Med. 368 (15), 1388–1397. doi:10.1056/NEJMoa1204471.

Weaverdyck, M.E., Lieberman, M.D., Parkinson, C., 2020. Tools of the trade. Multivoxel pattern analysis in fMRI: a practical introduction for social and affective neuroscientists. Soc. Cogn. Affect. Neurosci. 15 (4), 487–509. doi:10.1093/scan/nsaa057.

Woo, C.W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nat. Neurosci. 20 (3), 365–377. doi:10.1038/nn.4478.

Xie, J., Fang, P., Zhang, Z., Luo, R., Dai, B., 2021. Behavioral inhibition/activation systems and depression among females with substance use disorder: the mediating role of intolerance of uncertainty and anhedonia. Front. Psychiatry 12, 644882. doi:10.3389/fpsyt.2021.644882.

Xiu, J., Zhang, Q., Zhou, T., Zhou, T.T., Chen, Y., Hu, H., 2014. Visualizing an emotional valence map in the limbic forebrain by TAI-FISH. Nat. Neurosci. 17 (11), 1552–1559. doi:10.1038/nn.3813.

III Appendix

List of all publications

Klein, S., Kruse, O., Tapia León, I., van Oudenhove, L., van 't Hof, S. R., Klucken, T., Wager, T. D., & Stark, R. (2022). Cross-paradigm integration shows a common neural basis for aversive and appetitive conditioning. *NeuroImage*, *263*, 119594. https://doi.org/10.1016/j.neuroimage.2022.119594

Klein, S., Krikova, K., Antons, S., Brand, M., Klucken, T., & Stark, R. (2022). Reward Responsiveness, Learning and Valuation Implicated in Problematic Pornography Use – a Research Domain Criteria Perspective. *Current Addiction Reports*, *11*(288), 129. https://doi.org/10.1007/s40429-022-00423-w

Stark, R., Markert, C., Kruse, O., Walter, B., Strahler, J., & **Klein, S.** (2022). Individual cortisol response to acute stress influences neural processing of sexual cues. *Journal of Behavioral Addictions*, *14*(10), 1338. https://doi.org/10.1556/2006.2022.00037

van't Hof, S., van Oudenhove, L., Janssen, E., **Klein, S.**, Reddan, C., Kragel, P., Stark, R. & Wager, T. (2021). The Brain Activation-Based Sexual Image Classifier (BASIC): A Sensitive and Specific fMRI Activity Pattern for Sexual Image Processing. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhab397

Markert, C., **Klein, S.**, Strahler, J., Kruse, O. & Stark, R. (2021). Sexual incentive delay in the scanner: Sexual cue and reward processing, and links to problematic porn consumption and sexual motivation. *Journal of Behavioral Addictions*. https://doi.org/10.1556/2006.2021.00018

Klein, S., Kruse, O., Markert, C., Tapia León, I., Strahler, J., & Stark, R. (2020). Subjective reward value of visual sexual stimuli is coded in human striatum and orbitofrontal cortex. *Behavioral Brain Research*. https://doi.org/10.1016/j.bbr.2020.112792

Klein, S.*, Kruse, O.*, Tapia León, I., Stalder, T., Stark, R., Klucken, T. (2019) Increased neural reactivity to emotional pictures in men with high hair testosterone concentrations. *Social Cognitive and Affective Neuroscience*

Klucken, T., Kruse, O., **Klein, S.**, Kampa, M., Tapia León, I., & Stark, R. (2019). The relationship between neuroticism and appetitive conditioning. *Neurobiology of Learning and Memory*, *164*, 107068. https://doi.org/10.1016/j.nlm.2019.107068

Kruse, O., **Klein, S.**, Tapia León, I., Stark, R., & Klucken, T. (2020). Amygdala and nucleus accumbens involvement in appetitive extinction. *Human Brain Mapping.* https://doi.org/10.1002/hbm.24915

Stark, R., **Klein, S.**, Kruse, O., Weygandt, M., Leufgens, L. K., Schweckendiek, J., & Strahler, J. (2019). No Sex Difference Found: Cues of Sexual Stimuli Activate the Reward System in both Sexes. *Neuroscience*, *416*, 63–73. https://doi.org/10.1016/j.neuroscience.2019.07.049

*shared authorship

Statement regarding good scientific practice

I hereby declare that I have prepared the thesis at hand independently and without undue aid or the use of any resources other than indicated within the thesis. All parts of my thesis taken either verbatim or analogously from the published or unpublished works of or based on oral communications with others are indicated as such. Regarding all aspects of my scientific enquiries as they appear in my thesis, I have upheld the tenets of good scientific practice as laid out in the "Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" and complied with the precept of ethics, data protection and animal welfare. I declare that I have neither directly nor indirectly given monetary or any other valuable considera□ons to others in connec□on with the thesis at hand. I declare that I have not presented the thesis at hand, either in an identical or similar form, to an examination office or agency in Germany or any other country as part of any examination or degree. All materials from other sources as well as all works performed by others used or directly referenced within the thesis at hand have been indicated as such. In particular, all persons involved directly or indirectly in the development of the thesis at hand have been named. I agree with the screening of my thesis for plagiarism via offline or online detection-software.

Sanja Klein