UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Institut für Systemische Neurowissenschaften

Prof. Dr. Christian Büchel

Aversion processing following chronic opioid use with and without addiction

Dissertation

zur Erlangung des Doktorgrades PhD
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Caspar Jakob Geißler
aus Flensburg

Hamburg 2025

Betreuer:in / Gutachter:in der Dissertation:	**Prof. Dr. Christian Büchel**
Gutachter:in der Dissertation:	**Prof. Dr. Ingo Schäfer**
Vorsitz der Prüfungskomission:	**Prof. Dr. Christian Büchel**
Mitglied der Prüfungskomission:	**Prof. Dr. Ingo Schäfer**
Mitglied der Prüfungskomission:	**PD Dr. Regine Klinger**

Datum der mündlichen Prüfung: **12.09.2025**

Veröffentlicht mit Genehmigung der Medizinischen Fakultät der Universität Hamburg.

Table of contents

List of abbreviations

AUDIT	Alcohol Use Disorder Identification Test
BDI-II	Beck Depression Inventory II
BIDS	Brain Imaging Data Structure
BOLD	blood oxygen level dependent
CI	confidence interval
CNS	central nervous system
CS	conditioned stimulus
CSF	cerebro-spinal fluid
CR	conditioned response
CTQ	Childhood Trauma Questionnaire
CUDIT-R	Cannabis Use Disorder Identification Test Revised
dACC	dorsal anterior cingulate cortex
DASS-21	Depression-Anxiety-Stress Scale
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth edition
EDA	electro-dermal activity
EIMT	Emotional Intensity Morphing Task
EPI	echo-planar imaging
fMRI	functional magnetic resonance imaging
FTND	Fagerström Test for Nicotine Dependence
FoV	field of view
FWHM	full-width half-maximum
GABA	γ-aminobutyric acid
mOFC	medial orbitofrontal cortex
MPRAGE	magnetisation prepared rapid gradient echo
MRI	magnetic resonance imaging
NAcc	nucleus accumbens
OCD	obsessive-compulsive disorder
OCDUS	Obsessive-Compulsive Drug Use Scale
OIH	opioid-induced hyperalgesia
OMT	opioid maintenance therapy
OUD	opioid use disorder
PAG	periaqueductal grey
PCC	posterior cingulate cortex
PE	prediction error
PET	positron emission tomography
PIT	Pavlovian-to-Instrumental transfer
PNS	peripheral nervous system
POMI	Prescription Opioid Misuse Scale
preSMA	pre-supplementary motor area
PSS-10	Perceived Stress Scale
ROI	region of interest
S1	primary somatosensory cortex
SCL	skin conductance level
SCR	skin conductance response
SOWS	Short Opiate Withdrawal Scale
STAI-SF-S	Spielberger State Trait Anxiety Inventory – Short Form - state
TE	echo time
TR	recording time
US	unconditioned stimulus
VAS	visual analogue scale
vmPFC	ventromedial prefrontal cortex
VTA	ventral tegmental area

List of tables and figures

Opioid drugs are among the most ambiguous elements in the landscape of modern medicine. On the one hand, they are indispensable for treating acute and chronic pain, exemplified by the inclusion of opioid medications on the World Health Organisation's list of essential medicines (World Health Organization, 2023). For many people with chronic pain, opioids can provide at least short-term relief (Nury et al., 2022), which explains why a rising prevalence of chronic pain (Nahin et al., 2019, 2023) is accompanied by an increase in global opioid sales (Ju et al., 2022). On the other hand, repeated use of opioids - either obtained by prescription or illicitly - can have grave consequences, including opioid addiction, a compulsive pattern of continued opioid use despite significant distress and severe health consequences, including the risk of fatal overdose (Strang et al., 2020). In the present day, potential consequences of opioid use are nowhere as exemplified as in Northern America, where the United States finds itself in what has been called *opioid crisis* or *opioid overdose epidemic*. Since the year 2000, the number of people using opioids saw a steep increase and, dramatically, by 2021, the number of annual fatalities from opioid overdoses had increased fivefold compared to the start of the millennium (Centers for Disease Control and Prevention, 2024b), with over 100,000 U.S overdose deaths in 2023. At present, drug overdoses, most of which involve some opioid, constitute the leading cause of injury-related death in the United States (Centers for Disease Control and Prevention, 2024a). In Germany, around 166 000 people are estimated to be addicted to opioids (German Federal Ministry of Health, 2018) and more than 1 200 drug-induced deaths with opioid involvement were reported in 2021 (European Monitoring Centre for Drugs and Drug Addiction, 2023). Apart from fatal overdoses, opioid addiction can significantly impact physical and mental quality of life (Griffin et al., 2015) and has proven to be extremely difficult to overcome as indicated by frequent relapses even after prolonged bouts of abstinence (Hser et al., 2015).

However, despite the high number of individuals addicted to opioids, it is important to point out that only a minority of people who use opioids eventually develop addiction. When prescribed for the management of acute (e.g. post-operative) or chronic (e.g. cancer treatment) pain, the rate of opioid addiction was estimated to lie between 8 and 12% of patients (Vowles et al., 2015). Depending on the study design, diagnostic criteria, and data source, it might be even lower (Han et al., 2017; Higgins et al., 2018; Just et al., 2018; Kendler et al., 2023), implying that most pain patients maintain control over their opioid use. Among people who use opioids obtained illicitly (i.e. non-medically), rates of opioid addiction are somewhat higher – estimated to range from 13 to 25% (Becker et al., 2008; Elliott & Jones, 2019; Martins et al., 2017).

This goes to show that chronic opioid use does not necessarily lead to addiction. In fact, while some transition into addiction, most maintain control over their use, prompting the critical question of why

some people develop opioid addiction while others do not. Answering it carries tremendous importance in addressing two health care challenges simultaneously – providing adequate relief for an increasing number of people with chronic pain while minimising the risk of addiction that chronic opioid use brings.

To understand why people develop addiction to opioids, it is worthwhile to first take a step back and reflect on what opioids are, what people use them for, and how they exert their effects. "Opioid" is an umbrella term, coined to denote a group of substances – some illicit, others available on prescription - with "opium-like" properties. The name-giving opium – dried juice obtained from the seed capsule of the *papaver somniferum* poppy plant – has been used since Neolithic times for purposes of alleviating pain and treating insomnia (Brownstein, 1993). Its derivatives laudanum, morphine and diacetylmorphine (better known as heroin) were immensely popular and widely prescribed for pain relief until the mid-20th century. Today, various synthetic opioids are available for treating acute and chronic pain, each supposedly improved in its efficacy, safety profile, or abuse potential. What these substances have in common (and what is also reflected in their colloquial name "painkillers") is that they are primarily used to avoid unpleasant states, such as pain, stress or anxiety (Han et al., 2017). Critically, opioids can provide such avoidance because they change the perceived aversiveness of stimuli or subjective states. While alleviating physical pain is the most prominent effect of opioid-induced changes to aversion processing – mechanisms related to experiencing the disliking of a stimulus or state, usually accompanied by avoidance behaviour (American Psychological Association, 2018) - several other aspects are similarly impacted by opioids. For example, following acute opioid use, feelings of anxiety are reduced (Eikemo et al., 2023), physiological stress reactions are attenuated (Bershad et al., 2015, 2018; Massaccesi et al., 2022), negative effects of social rejection are dampened (Bershad et al., 2015, 2016), and recognition of fearful or angry facial expressions is reduced (Ipser et al., 2013; Løseth et al., 2018). It is crucial to point out that even though opioids can increase feelings of pleasure and induce euphoria in some, this is not the case for the majority of individuals (Eikemo et al., 2023), and avoidance of aversive states has been repeatedly reported to be the principal motivation driving opioid use (Han et al., 2017; McHugh et al., 2022; Weiss et al., 2014). Indeed, Comer et al. (2010) coined the phrase "negative euphoria" to describe that opioids elicit pleasant feelings only because aversion was reduced.[1]

Pharmacologically, opioids act by binding on a certain group of receptors widely distributed throughout the human body (Al-Hasani & Bruchas, 2011; Peng et al., 2012). These opioid receptors (mainly μ-, δ-, κ- and ORL1-receptors) play a key role in modulating several vital systems, such as the

[1] Poignantly, when reflecting upon his heroin addiction, the author William Burroughs (1914-1997) contemplated a similar idea: "I have experienced the agonising deprivation of junk [heroin] sickness, and the pleasure of relief when junk-thirsty cells drank from the needle. Perhaps all pleasure is relief." (Burroughs, 2008, p. 41)

central and peripheral nervous system (CNS and PNS), the cardio-vascular system and the gastro-intestinal system. Opioid drugs are structurally very similar to opioid receptors' natural ligands, endogenous opioids, a class of peptides synthesised in the human brain. The release of endogenous opioids determines the opioidergic tone and regulates the activity of brain networks with a high opioid receptor density (Nummenmaa & Tuominen, 2018). Experimental manipulation of the opioidergic tone demonstrates that the opioid system modulates how pleasant or unpleasant a stimulus or state is perceived to be. When the opioidergic tone is reduced by blocking opioid receptor function through the administration of an antagonist, appetitive stimuli are perceived as less rewarding (Meier et al., 2021). Conversely, following antagonist treatment, fear learning is increased and more sustained (Eippert et al., 2008), money loss is perceived as more unpleasant (Petrovic et al., 2008), and while opioid antagonists do not necessarily worsen the experience of pain, they appear to fine-tune it (Eikemo et al., 2021). Opioids[2], on the other hand, increase opioidergic tone by binding to opioid receptors and mimicking the effects of endogenous ligands (Al-Hasani & Bruchas, 2011), which then also results in altered functioning of brain areas rich in opioid receptors (Becerra et al., 2006; Kantonen et al., 2020; Leppä et al., 2006).

Repeatedly increasing the tone of the opioid system by way of opioid use, however, introduces changes to its functioning. Chronic opioid use (i.e. regular use for at least three months; Chou, 2015) can lead to structural and functional alterations of μ-receptors, such as desensitisation and accelerated receptor internalisation, which in turn prompt changes in the behaviour of pre-synaptic neurons (e.g. reduced transmitter release probability), ultimately resulting in the dysregulation of entire neural systems modulated by endogenous opioids (Christie, 2008). One indicator of an altered opioid system is the build-up of tolerance: Over time, a dose that initially caused a certain effect will cease to have the same effect – to reach the same effect again, a higher dose needs to be administered (Christie, 2008). Underlying this is an allostatic state introduced by chronic opioid use: a disruption to the "natural flow" (homeostatic state) of neural signalling (Cahill et al., 2016; Koob, 2020) where normal functioning is only maintained when opioids are continuously administered. Once the opioid is metabolised and loses its effect, withdrawal sets in. In contrast to the opioid-induced effects of reduced pain perception, relaxation, and anxiolysis, this withdrawal state is characterised by sensitization to pain, dysphoria, irritability, and pronounced bouts of anxiety (Srivastava et al., 2020). This is often complemented by somatic withdrawal symptoms such as tachycardia, diarrhoea, and insomnia. Withdrawal symptoms last until the opioid system has regained homeostasis, which can take from a few days up to two weeks after last use (Srivastava et al., 2020). Critically, withdrawal ceases

[2] Formally, all substances (endogenously produced or exogenously administered) that bind to opioid receptors are considered opioids. As this thesis concerns the chronic use of opioid drugs, i.e. exogenously administered drugs which primarily act as agonistics on μ-receptors, from here on I will use the term "opioid" for those substances, unless stated otherwise.

instantly once a sufficiently high opioid dose is taken, rendering avoidance of withdrawal a motivation for continued opioid use (Koob, 2020). This interplay between opioid use, tolerance, and withdrawal represents a physical dependence on opioids, which represents a normal response to chronic use, where the extent differs between individuals (e.g. Ware & Dunn, 2023). Importantly, dependence is phenomenologically distinct from addiction and is not sufficient for a diagnosis of opioid addiction (Ballantyne et al., 2019).

Given that chronic use leads to adaptations of the opioid system and the key role of the opioid system for aversion processing, it is reasonable to expect that chronic opioid use also alters aversion processing. First evidence indeed supports this assumption: Patients with chronic use due to either addiction or chronic pain report higher pain sensitivity (opioid-induced hyperalgesia; Higgins et al., 2019; Trøstheim & Eikemo, 2024), stronger stressful responses to social rejection (Kroll et al., 2019), higher feelings of social exclusion (Bach et al., 2019) as well as increased emotional responses to negative affective images and fearful facial expressions (Aguilar De Arcos et al., 2008; Schmidt et al., 2014). These findings remarkably contrast the effects of acute opioids in healthy volunteers, where pain sensitivity was reduced, reactivity to negative emotions attenuated, and effects of social rejection and social stress were dampened (see above) and are consistent with a decreased opioidergic tone following chronic opioid use, possibly due to reduced pre-synaptic receptor availability.

To summarise, an increasing number of people use opioids to avoid aversive states, most prominently physical pain, which can be explained by opioids' effects on brain areas implicated in aversion processing. When used chronically, the opioid system changes, which might in turn cause changes in aversion processing. To this date, it is still unclear which aspects of aversion processing are altered following chronic opioid use. It is also unknown whether individual differences in aversion processing contribute to the development and maintenance of opioid addiction. Characterising how aversion processing is altered following chronic opioid use and its potential specificity to opioid addiction will help researchers better understand how opioid addiction develops. More importantly, it will help practitioners and patients make more informed decisions about prescribing opioids, maximising their therapeutic potential while simultaneously minimising the risk of addiction.

In this thesis, I seek to characterise aversion processing following chronic opioid use in patients with and without addiction. I will first describe the role of the opioid system in various aspects of aversion processing, before summarising existing evidence for a relationship between opioid addiction and aversion processing (Chapter 2). I will conclude this chapter by stating the hypotheses for this thesis. In Chapter 3, I will present the general methodology applied to test these hypotheses. In the chapters to follow (Chapters 4 through 8), I will investigate different aspects of aversion processing following

chronic opioid use with and without addiction. Finally, in Chapter 9, I will discuss my findings in more depth and conclude with implications for future research as well as for clinical practice.

Chapter 2 - Background

In the following, I will review the existing literature in regards to a proposed bidirectional relationship between opioids and aversion processing. On the one hand, endogenous opioids are involved in aversion processing and acute opioid use alters aversion processing; on the other hand, when opioids are used chronically, inter-individual differences in aversion processing might contribute to who develops opioid addiction and who does not.

2.1 - Aversion processing and the opioid system

Aversion processing refers to a group of psychological mechanisms related to responses towards aversive stimuli or states, i.e. stimuli or states that are assigned a negative value and that are therefore disliked and undesirable[3] (American Psychological Association, 2018). It encompasses a multitude of mechanisms, spanning from the recognition of aversive emotions to the ability to learn from aversive experiences to higher-order processes, such as regulatory control over avoidance behaviour. Understanding the role of the opioid system in these processes constitutes a first step toward understanding how chronic opioid use might alter aversion processing and potentially contribute to the development of opioid addiction in some, but not in other, individuals. Therefore, in the following, I will introduce key aspects of aversion processing and review their neural underpinnings with a focus on the opioid system.

Recognition of aversive emotions

Recognising another person's emotional state is a critical social skill that allows us to adapt our behaviour towards others and their current needs. As such, emotion recognition is closely tied to empathy (Besel & Yuille, 2010) and social intelligence (Austin, 2004). Beyond its importance in the social domain, impaired recognition of others' emotions is associated with the ability to identify and describe one's feelings (alexithymia; Sifneos, 1973). Especially difficulties in recognising fear and anger are related to impairments in perspective-taking (Trubanova et al., 2015) and theory of mind (Corden et al., 2006) as well as increased anti-social behaviour (Marsh & Blair, 2008). The ability to recognise fear, anger, and other aversive emotions in other's faces therefore represents a crucial aspect of aversion processing, with impairments potentially detrimental to social relationships and emotional well-being.

[3] Its antonym, reward processing, covers responses related to appetitive stimuli or states that are assigned positive values.

Recognising emotions in facial expressions is a multi-layered process with complex underlying neurobiology. For successful emotion recognition, certain facial actions (face action units, e.g. raised eyebrows, lips stretched toward ears) are detected, combined, and used as cues, in reaction to which attention is directed to particularly relevant facial features (such as the eye region). Simultaneously, information about emotional categories is retrieved from memory and the observed facial actions are simulated (facial mimicry), leading to the correct classification of the specific emotion in question (Spunt & Adolphs, 2019). These sub-processes rely on various brain regions: the superior temporal sulcus serves the initial recognition of facial actions (Srinivasan et al., 2016), the amygdala shifts attention to emotionally relevant features and recruits additional cortical brain areas (Diano et al., 2017; Gamer & Büchel, 2009), and the somatosensory cortex simulates embodied emotions (Kragel & LaBar, 2015). Perturbations to these brain areas significantly impair the ability to recognise negative emotions. Lesions to the amygdala in particular reduce accuracy in recognising fear (Adolphs et al., 1994) and anger (Adolphs & Tranel, 2003), which can be explained by reduced attention towards information-rich facial features (Adolphs et al., 2005).

Experimental studies have demonstrated the involvement of the opioid system in recognising aversive emotions in faces. Administration of the μ-agonist and partial κ-antagonist buprenorphine leads to impaired fear recognition compared to placebo (Ipser et al., 2013). Likewise, buprenorphine also reduces attention to fearful faces (Bershad et al., 2016), which is thought to underlie the impaired fear recognition. Administration of the μ-agonist morphine reduced the perception of anger in faces with neutral or subtle facial expressions (Løseth et al., 2018), while oxycodone did not alter subjective or neural responses to emotional facial expressions (Wardle et al., 2014). Additionally, the non-selective opioid antagonist naltrexone delayed recognition of sadness and fear in others' faces (Wardle et al., 2016), which could be explained by reduced visual exploration and attention to the eye region following naltrexone administration (Chelnokova et al., 2016).

These studies demonstrate that the opioid system is involved in the recognition of negative emotions. Directing attention towards relevant facial features for identifying fearful (and potentially angry) faces could to be a critical mechanism of emotion recognition that is modulated by opioids. The fact that the key neural substrate for emotionally directed attention, the amygdala, is dense in μ- and κ-receptors (Pfeiffer et al., 1982) underscores this notion.

Fear conditioning and safety learning

Beyond mere identification of whether a stimulus is aversive or not, there is a need to anticipate whether an aversive event will occur, providing oneself with the possibility to react accordingly. Environmental cues that have been paired with aversive events are learned by association to elicit fear,

an affective response to anticipated threat (Mobbs et al., 2019), prompting a defensive reaction before the actual events occur. In experimental psychology, fear conditioning (also called aversive Pavlovian conditioning) describes the mechanism through which the presentation of a previously innocuous cue (conditioned stimulus; CS) elicits a fear response after being paired with an aversive stimulus (unconditioned stimulus (US); Fullana et al., 2016). Safety learning, conversely, describes the mechanism through which a cue elicits a safety response after being repeatedly followed by an absence of the US (Harrison et al., 2017). To facilitate an adaptive emotional response to the presence and absence of threat, both mechanisms need to be balanced, with fear responses expressed towards threatening cues and safety responses towards non-threatening cues. A disruption to this balance, such as expressing fear responses towards safety cues, can become highly debilitating and has been reported in patients with anxiety disorders and post-traumatic stress disorder (Mahan & Ressler, 2012; Pittig, Treanor, et al., 2018).

In the brain, learning of fear and safety are encoded in widespread networks. A meta-analysis of fMRI fear conditioning studies showed increased activation in a so-called *fear network* (including anterior insula and dorsal anterior cingulate cortex; dACC) during CS presentation in the acquisition phase, i.e. when the CS is followed by the US (Fullana et al., 2016). Simultaneously, activation of a *safety network* (including the ventromedial prefrontal cortex; vmPFC) is suppressed. When the US is no longer present, fear network activation decreases over time (extinction phase), suggesting that the network tracks the aversive value of the CS (Fullana et al., 2018). Presentation of a safety cue activates the vmPFC, which leads to an inhibition of the fear response (Harrison et al., 2017). This demonstrates that distinct brain networks underlie the adaptive learning of fear and safety information about previously neutral context cues.

The opioid system has been implicated in fear conditioning through a line of animal studies, suggesting that fear and safety learning rely on opioid-dependent prediction error signals in multiple brain regions including the periaqueductal grey (PAG) and the amygdala (McNally et al., 2011). In humans, one study found that the administration of the non-selective opioid antagonist naloxone led to more sustained activation of the amygdala and enhanced acquisition of fear responses (Eippert et al., 2008). In accordance with animal research, this suggests that endogenous opioids inhibit fear responses after initial learning. Blocking opioid receptors disrupts this process, underscoring the role of the opioid system in efficient emotional learning and adaptive responses to aversive stimuli.

Instrumental avoidance learning

Instrumental avoidance learning presents another aversive learning mechanism, related but yet distinct from fear conditioning, that conveys learning from past aversive experiences. Instrumental

learning in general relies on the assumption that future behaviour is guided by the consequences of past behaviour: if a behaviour has led to a positive outcome (positive reinforcement) or the avoidance of a negative outcome (negative reinforcement), it should be carried out with higher probability or frequency in the future (Sutton & Barto, 2018). In instrumental avoidance learning, a behaviour that leads to an aversive outcome should be reduced or ceased entirely, while competing behaviours that led to the omission of that outcome should be selected more frequently. This provides an opportunity to avoid harm flexibly and adaptively: the more information about the relationship between a behaviour and its outcome is learned, the easier it is to avoid aversive outcomes by choosing correctly.

In the brain, instrumental avoidance learning relies on a fronto-striatal circuit that includes, among other structures, the anterior insula, the caudate nucleus in the dorsal striatum, and the medial orbitofrontal cortex (mOFC; Kim et al., 2006; Pessiglione et al., 2006). An unexpected aversive outcome (e.g. money loss) leads to activation of the anterior insula, where the aversive prediction error (difference between expected and actual outcome) is encoded and paired with the behaviour that has led to the outcome. Based on the expected outcome values of several competing behavioural options, the behaviour associated with the highest expected outcome value is selected in the caudate (Pessiglione et al., 2006). If the selected behaviour leads to the avoidance of an aversive outcome, the mOFC is activated, which putatively updates the expected values of the selected behaviour (Kim et al., 2006; Palminteri et al., 2012), leading to an accumulation of evidence for or against choosing behavioural options over time.

While instrumental learning in generally is thought to be dependent on dopaminergic innervation of the striatum (Pessiglione et al., 2006), instrumental avoidance could also modulated by the opioid system. In a recent study, participants learned to avoid a painful stimulus, and while opioid receptor blockage failed to alter overt avoidance behaviour, a computational parameter indicating successful avoidance learning was increased (Jepma et al., 2022). One explanation for this potentially improved avoidance learning could be that endogenous opioids attenuate aversive prediction errors in the insula; blocking opioid transmission would potentiate prediction errors and facilitate avoidance learning. Findings that naloxone increased insula activation in response to aversive outcomes support this idea (Petrovic et al., 2008). Conversely, opioid agonists reduce insula activation following aversive stimuli (e.g. Wise, 2002) and potentially minimise aversive prediction errors. This could impair learning, as accurate tracking of aversive prediction errors is a central requirement for flexible avoidance. However, insula activity was not altered following naltrexone in the above mentioned study by Jepma et al. (2022), leaving the role of opioids in instrumental avoidance learning to be somewhat unclear.

Pavlovian-To-Instrumental transfer

Although fear conditioning and instrumental avoidance learning represent two distinct mechanisms through which affect and behaviour are shaped by aversive events, they are by no means independent from each other. Pavlovian-to-Instrumental transfer (PIT) denotes a process through which environmental cues that have acquired affective or motivational value by way of Pavlovian conditioning can exert influence on instrumental behaviour (Holmes et al., 2010). It can either be general (i.e. the outcome from instrumental learning matches the Pavlovian US) or specific (i.e. the outcome from instrumental learning is unrelated to the Pavlovian US; Holmes et al., 2010). PIT has been well documented for the appetitive domain, where the presentation of reward-associated background cues increases approach behaviour acquired through instrumental learning (Talmi et al., 2008). More recently, studies found a corresponding effect for the aversive domain where the presentation of cues associated with aversive outcomes increased avoidance behaviour (Garofalo & Robbins, 2017; Lewis et al., 2013; Xia et al., 2019). By and large, PIT is an adaptive mechanism that can increase context-specific readiness for learned behaviours, marked by shorter reaction times and increased vigour in experimental settings (Garofalo & Robbins, 2017; Huys et al., 2011); however, when PIT is disproportionally strong, environmental cues can bias behaviour to a point where a person performs approach or avoidance response even though this is not the optimal behaviour based on their (instrumental) learning experience. For example, the smell of a hamburger (paired with a reward - eating the burger - in the past) might elicit a strong desire for high-caloric food when choosing from a menu, biasing choice towards this option even though one was supposed to be on a diet (specific appetitive PIT). On the contrary, after receiving a call from your boss that has often criticised you harshly in the past, you ride your bike incredibly passively and carefully as the cue previously paired with punishment introduced an avoidance bias into unrelated behaviour (general aversive PIT).

Aversive PIT is enabled by the shared underlying neurobiology of fear conditioning and instrumental avoidance learning. Within the cortico-striatal circuits involved in instrumental avoidance learning, insula activation encodes aversive prediction errors which underlie the updating of aversive values, critical when selecting whether or not to carry out an avoidance behaviour. During PIT, insula activation is increased when the aversive CS is presented (Lewis et al., 2013), consistent with its role as part of the fear network identified in conditioning studies (Fullana et al., 2016). As a result, a higher aversive value is parsed through to the striatum, which biases decision-making in favour of avoidance. Additionally, while instrumental avoidance depends on the caudate for response selection (Pessiglione et al., 2006), PIT engages the putamen, a different sub-region of the dorsal striatum (Lewis et al., 2013) implicated in automatic responding rather than deliberate action selection (Patterson & Knowlton,

2018). Taken together, more reflexive, cue-driven responses and increased aversive outcome values could provide the neural basis for aversive PIT.

Given the role of opioids in fear conditioning and instrumental avoidance learning, it would be reasonable to expect opioidergic modulation of aversive PIT. Enhanced fear responses to an aversive CS (Eippert et al., 2008) and higher insula activation to aversive outcomes (Petrovic et al., 2008) following opioid receptor blockage could presumably lead to an increased PIT effect. This would tie in with the findings of increased aversive PIT in patients with major depression (Nord et al., 2018), a disorder characterised by low opioidergic tone (M. Peciña et al., 2019). One study tested opioidergic modulation of appetitive PIT and found that PIT-induced increases in instrumental responding for snack rewards were abolished following naltrexone administration (Weber et al., 2016). This would fit the idea of decreased appetitive values after opioid receptor blockage, which could in turn increase the CS aversive value. Nevertheless, whether a similar manipulation would indeed lead to a stronger aversive PIT signal remains unclear.

Avoidance habits

Avoidance behaviour acquired via instrumental learning at first is goal-directed, meaning a behaviour is selected based on the outcome it is expected to produce in a given context. As a consequence, if the outcome changes, so does the probability of selecting the behaviour, which is crucial for flexibly adapting behaviour to changing environments. However, when repeated extensively, behaviour can gradually shift into a habit (Dickinson, 1985). Habits are carried out reflexively in response to a stimulus, without considering the outcome of the behaviour. This makes habitual avoidance fast and low on cognitive resources, but also inflexible to changing outcome values. The trade-off between computational costs and flexibility renders goal-directed and habitual behaviour suitable under different circumstances: Habits can be safely performed in routine settings; however, once relevant features undergo change, goal-directed behaviour is required. The ability to switch between the two systems underlies the behavioural balance that is necessary to successfully navigate complex environments (Bouton, 2021). When this balance gets disrupted, either transiently due to stress or persistently due to an over-reliance on the habit system (e.g. as part of a psychiatric disorder), avoidance habits can dominate behaviour, as can be seen, for example, in patients with obsessive-compulsive disorder (OCD), who might compulsively check whether they left on the oven to avoid their kitchen catching fire, despite knowing that their behaviour does not actually contribute to accomplishing this goal (Gillan et al., 2014).

Neuroimaging has identified separate, yet interacting, neural substrates for the goal-directed and the habit system, implicating the vmPFC and caudate nucleus in goal-directed actions and the posterior

putamen and supplementary motor area in habitual responses (De Wit, Watson, et al., 2012). In animal studies, the shift from goal-directed to habitual behaviour is mirrored by a ventral-to-dorsal shift in striatal activity (Gremel & Costa, 2013), which corresponds to the anatomy of caudate and putamen findings in humans. Activation of the insula, previously mentioned for its important role in instrumental avoidance, decreases over time when avoidance behaviour transitions into a habit; strength of avoidance habits was predicted by reduced activity in the inferior parietal lobe, a brain region implicated in outcome anticipation (Zwosta et al., 2018). This shows that when goal-directed behaviour is extensively trained, reliance on outcomes decreases, promoting performance of outcome-insensitive habits. Maladaptive overreliance on habits has been associated with altered functioning of brain regions underpinning the goal-directed system, such as reduced vmPFC sensitivity to outcome values following administration of stress hormones (Schwabe et al., 2012) as well as caudate and vmPFC hyperactivation in patients with OCD (Gillan et al., 2015).

Several neurotransmitter systems have been implicated in the shift from goal-directed to habitual behaviour, including the dopamine (De Wit, Standing, et al., 2012), serotonin (Worbe et al., 2015), and noradrenaline system (Schwabe et al., 2012). Relatively little is known about the role of opioids in habit formation. One study used positron emission tomography (PET) to determine the binding potential for an μ-opioid tracer and correlated the findings with behaviour in a task measuring model-free and model-based learning, which can be understood as computational markers for habits and goal-directed behaviour (Voon et al., 2020). Results showed that binding potential in the nucleus accumbens (NAcc), a sub-region of the ventral striatum, was positively correlated with goal-directed avoidance behaviour and negatively correlated with habitual avoidance. High binding potential indicates lower NAcc opioid tone, which could enhance the impact of aversive outcomes and therefore facilitate goal-directed instrumental learning (similar to the effects of an opioid antagonist; Jepma et al., 2022; Petrovic et al., 2008). The observed low endogenous opioid tone could also generally reduce the acquisition of goal-directed behaviour, which would be consistent with an animal study in which naloxone during instrumental learning rendered the behaviour of rats less sensitive to changes in food value, i.e. more habitual (Wassum et al., 2009). However, in the study by Voon et al. (2020), the correlation between opioidergic tone and behaviour was found only for avoiding money losses, not for gaining money, supporting the role of opioids as specific to aversion in the interplay between goal-directed and habitual behaviour.

Taken together, while the opioid system has been implicated in many facets of aversion processing, its role has been better characterised in some aspects compared to others. Previous literature implicates endogenous opioids in processing aversive social information such as angry or fearful facial expressions and in aversive learning, while opioidergic involvement in regulatory control over avoidance behaviour

is less clear. As chronic opioid use alters the opioid system through various mechanisms (Christie, 2008) and coincidently changes the structure of key brain regions related to aversion processing (Lin et al., 2016; Younger et al., 2011), it would be plausible that aspects of aversion processing modulated by opioidergic tone would also be altered following chronic use.

2.2 - Aversion processing in opioid addiction

While the majority of patients with chronic opioid use maintains control over their opioid use (Vowles et al., 2015), some find themselves unable to stop use once negative consequences outweigh the positive effects, a key criterion indicative of opioid addiction (Strang et al., 2020). It remains largely unclear why some patients develop opioid addiction and how inter-individual differences in aversion processing contribute to this development. In the following sections, I will first characterise opioid addiction and its underlying neurobiology to then characterise the existing knowledge regarding the relationship between aversion processing and opioid addiction.

Opioid addiction

Opioid addiction (classified as severe Opioid Use Disorder; OUD) in the most recent Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013; see Table 1) is a chronically relapsing disorder defined by persisting opioid use despite significant negative consequences (Strang et al., 2020). It is characterised by a loss of control over opioid use, a compulsion to take opioids, and the onset of withdrawal when opioid use is stopped (Strang et al., 2020). Patients with opioid addiction find themselves in a vicious cycle of opioid use, highly aversive opponent processes once the opioid effect wears off (pronounced irritability, anhedonia, high pain sensitivity, accompanied by somatic symptoms), craving (i.e. strong desire to use the drug again), and subsequent opioid use (Koob & Volkow, 2016). In addition, opioid addiction is related to long-lasting changes in hedonic processing, characterised by low mood, heightened stress sensitivity, and anxiety (Koob, 2020). Opioid addiction is considered a chronic disorder since intense craving and relapse can arise even after prolonged abstinence when stress is high or upon encountering cues related to opioid use (Fatseas et al., 2011) so that patients often relapse after years without opioid use.

Harmful consequences of opioid addiction substantially depend on the individual and the opioid in question. Because of the development of tolerance, many patients either increase their opioid dose or switch to a more potent (or more readily available) opioid, often heroin (Cicero et al., 2014). The use of high-potency opioids like heroin increases the risk of a fatal overdose due to respiratory depression, which is further potentiated by combined use of opioids with other sedatives (e.g. alcohol,

benzodiazepines; White & Irvine, 1999). Another important factor regarding opioid-associated harm is the route of administration, with intravenous injections (which some patients prefer due to rapid onset of opioid effects and higher bioavailability) facilitating the contraction of blood-borne diseases through needle-sharing, such as HIV or Hepatitis C (S.-C. Wang & Maher, 2019). Chronic opioid use also impacts the healthy functioning of endocrine and immune systems, leading to hypogonadism, increased chronic stress, and weaker immune responses in many patients (De Vries et al., 2020; Malafoglia et al., 2022). Also, possession of opioid drugs is restricted or illegal in most parts of the world, prompting criminal charges against patients with many serving prison sentences for possession of opioids or criminality associated with financing opioid use (Winkelman et al., 2018).

Table 1. Diagnostic criteria of Opioid Use Disorder according to DSM-5. Diagnosis of opioid addiction (severe OUD) requires the fulfilment of a minimum of 6 of the 11 criteria.

1.	Taking the opioid in larger amounts and for longer than intended
2.	Wanting to cut down or quit but not being able to do so
3.	Spending a lot of time obtaining the opioid
4.	Craving or a strong desire to use opioids
5.	Repeatedly unable to carry out major obligations at work, school or home due to opioid use
6.	Continued use despite persistent or recurring social or interpersonal problems caused or made worse by opioid use
7.	Stopping or reducing important social, occupational or recreational activities due to opioid use
8.	Recurrent use of opioids in physically hazardous situations
9.	Consistent use of opioids despite acknowledgment of persistent or recurrent physical or psychological difficulties from using opioids
10.	Tolerance as defined by either a need for markedly increased amounts to achieve intoxication or desired effect or markedly diminished effect with continued use of the same amount[4]
11.	Withdrawal manifesting as either characteristic syndrome or the substance is used to avoid withdrawal

Treatment approaches for opioid addiction aim to break the cycle of addiction as well as alleviate health detriments associated with use. This primarily entails the prescription of medications targeted at suppressing withdrawal (opioid maintenance therapy; OMT), which are mainly longer-acting opioid agonists such as methadone and buprenorphine as well as oral morphine and heroin in some countries (Haasen et al., 2007; Strang et al., 2015; Volkow et al., 2019). Following opioid detoxification, OMT is generally well tolerated, reduces illicit opioid use and mortality, restores immune function, and often serves as an entry point into additional addiction treatment services (Gibson et al., 2008; Sacerdote et al., 2008; Soyka et al., 2017). Nevertheless, a substantial number of patients in OMT still experiences relapse (Hser et al., 2015). So-called harm reduction approaches include free needle exchange programs, treatment of infections associated with intravenous use, and disposal of naloxone nasal spray to reverse opioid overdoses (Wakeman, 2019). Psychosocial treatment approaches can be

[4] In contrast to previous versions, DSM-5 acknowledges that criteria 10 and 11 (tolerance and withdrawal) are not considered to be met for patients taking opioids solely under medical supervision, reiterating that physical dependence presents a non-pathological adaptation to chronic use (Ballantyne et al., 2019).

adjunct to pharmacological therapy and aim to increase motivation for non-drug behaviours, reduce craving and prevent stress-induced relapse (Dugosh et al., 2016).

It is crucial and bears repeating that, while continued opioid use is necessary for developing opioid addiction, it is not sufficient. The fact that some individuals with chronic use transition into addiction while others do not suggests an interplay between vulnerabilities and protective factors. In its aetiology, opioid addiction is considered a multi-faceted biopsychosocial disorder, with several cumulative and interacting risk factors, including young age at opioid use onset, male gender, low socio-economic background and education level, presence of affective disorders, family history of addiction and childhood abuse or neglect (Strang et al., 2020; Webster, 2017). It is unclear through which neurobiological or psychological processes these risk factors exert their effects on the development of opioid addiction. While some risk factors, such as adverse childhood experiences and psychiatric symptoms, have been linked to altered aversion processing (e.g. Ruge et al., 2024) or opioids effects on aversion processing (Carlyle et al., 2021), the link between aversion processing and risk factors for opioid addiction has not been systematically investigated.

Neurobiology of opioid addiction

Generally, the neural mechanisms underlying drug addiction are broken down into three stages: the binge/intoxication stage, where drugs are used and exert their effects; the withdrawal/negative affect stage when drugs are not available; and the preoccupation/anticipation stage, characterised by preoccupation with and a strong desire to use the drug (Koob & Volkow, 2016). From initial drug use to addiction, the stages continually repeat and intensify with each iteration. They subsequently lead to neuroadaptations (i.e. changes in brain structure and function), with each stage of the addiction cycle associated with changes in different brain regions. Increased salience and motivational properties of drug-related cues (such as the local pub, pills or syringes) during the binge/intoxication phase has been linked to altered function of the brain's reward system, including the ventral tegmental area (VTA) and NAcc (Robinson & Berridge, 1993; Zilverstand et al., 2018). Loss of regulatory control over drug use during this phase has been linked to the ventral-to-dorsal shift in striatal activity indicative of habit formation (Everitt & Robbins, 2016). Neuroadaptations in the brain's stress system (sometimes called the anti-reward system) centring around the amygdala have been associated with the long-lasting aversive emotional states (hyperkatifeia) and stress sensitivity of the withdrawal/negative affect stage (Koob, 2013, 2020). During the preoccupation/anticipation stage, the inability to maintain top-down control over drug use has been linked to a dysfunction of inhibitory cortico-striatal loops between the striatum and prefrontal cortex (Goldstein & Volkow, 2011), while

the intensity of craving is at least partly reliant on a network connecting the insula and the dorsal striatum (Naqvi & Bechara, 2009).

The neuroadaptations postulated in this model are mostly based on pre-clinical studies and not specific to opioid addiction; however, evidence for corresponding effects in patients with opioid addiction exists. Meta-analyses of studies on opioid-addicted patients found decreased grey matter volume and impaired white matter integrity in medial frontal brain regions (Wollman et al., 2015, 2017), in line with impaired top-down control. Additionally, neuroimaging studies have identified smaller volumes of limbic and striatal brain regions, such as the NAcc, putamen, and amygdala in opioid-addicted patients in OMT compared to healthy controls (Ceceli, King, et al., 2023; Schmidt et al., 2021; Seifert et al., 2015). These structural changes correspond to changes in neural function. Duration of opioid use was correlated with stronger decreases in functional connectivity between the insula and the putamen, whereas connectivity between NAcc and mOFC was increased (Upadhyay et al., 2010), indicating a re-organisation of cortico-striatal pathways in opioid addiction. When processing drug-related images, a mOFC, cingulate cortex, amygdala, and caudate nucleus among other regions were hyper-activated in patients compared to healthy controls (Dejoie et al., 2024), while higher NAcc reactivity to heroin-related pictures was fount predictive of higher risk of relapse in another study (Q. Li et al., 2015).

Within the three-stage model of addiction, initial drug use is typically understood to be motivated by rewarding drug effects brought on by phasic bursts of dopamine in VTA-NAcc networks (Koob & Volkow, 2016). Over time, these effects diminish as tolerance develops and the so-called hedonic set-point shifts (R. A. Wise & Koob, 2014), inducing opponent processes characterised by negative affective states. As a consequence, the motivation for drug use is thought to switch from approaching reward to avoiding aversion. Correspondingly, the underlying behavioural mechanism gradually transitions from positive to negative reinforcement. However, since it is becoming increasingly clear that most people use opioids to avoid negative affective states (Han et al., 2017), recent accounts acknowledge that in the case of opioids specifically, initial use can already be driven by the motivation to avoid chronic pain or self-medicate other types of emotional distress (Koob, 2020). In this case, individuals enter the drug use cycle at the negative affect stage, where the anticipation of pain relief motivates opioid use and positive reinforcement plays a subordinate role (supported by a smaller role of the dopamine system in opioid addiction compared to other drugs of abuse; Badiani et al., 2011; Daglish et al., 2008). Over time, avoidance of the initial aversive state is overtaken by the motivation to avoid withdrawal. This characterises the development of opioid addiction as a progressively escalating avoidance behaviour, which gradually transitions from goal-directed to compulsive, stimulus-driven avoidance.

The bidirectional relationship between aversion processing and opioid addiction

Many brain regions with altered structure and function in opioid-addicted patients have also been implicated in aversion processing, such as the vmPFC/mOFC, insula, amygdala, and striatum. This is not surprising given that aversion processing depends on opioidergic modulation and can acutely be altered by opioids. Lower opioidergic tone, indicated by reduced μ-receptor availability, has been reported in patients in OMT (Kling et al., 2000; Zubieta, 2000) and altered aversion processing is well documented in opioid addiction. For example, opioid-addicted patients in OMT show increased sensitivity to pain (Trøstheim & Eikemo, 2024), increased emotional responses to aversive pictures (Aguilar De Arcos et al., 2008; Gerra et al., 2003), and stronger responses towards social rejection and exclusion (Bach et al., 2019; Kroll et al., 2019). In one study, patients in OMT showed significantly higher amygdala responses to pictures of fearful faces, which were normalised to the degree of healthy controls after injection of heroin (Schmidt et al., 2014). Opioid-addicted patients also showed impaired fear conditioning in one study (Basden et al., 2016) and decreased neural responses to loss-avoidance in the ventral striatum in another (Gradin et al., 2014).

The studies referenced above provide evidence that some aspects of aversion processing are altered in patients with opioid addiction. Other key aspects of aversion processing related to regulatory control over avoidance behaviour, namely habit formation and disproportionate PIT, are considered crucial to the development of addiction. The habit theory of addiction posits that an imbalance between goal-directed and habitual behaviour serves as a primary building block for compulsive drug use (Everitt & Robbins, 2005, 2016). While initial drug use is goal-directed, continued use causes a shift toward habitual drug use, where use occurs automatically in certain contexts. Evidence from rodent research suggests that following extensive training, responding for drug administration becomes a habit and a vulnerable subgroup of animals develops compulsive drug seeking (i.e. continued cue-elicited drug seeking even when no more drugs could be obtained and drug-seeking behaviour itself was punished; Everitt et al., 2008). According to this theory, an imbalance between habits and goal-directed behaviour[5] combined with decreased top-down inhibitory control could lead to compulsive drug use. While this has not been investigated in opioid addiction, an impaired balance between goal-directed and habitual behaviour has been demonstrated in patients addicted to alcohol (Sjoerds et al., 2013) and cocaine (Ersche et al., 2016). Note that these two studies found increased appetitive (i.e. reward-based) habits; Ersche et al. (2016) additionally tested for avoidance habits in patients with cocaine

[5] Whether this imbalance is caused by excessive habit formation, impaired goal-directed control or by defunct arbitration between the two systems is still under debate (see Vandaele & Janak, 2018).

addiction, who did not differ from controls in this regard. This points to a valence-specific dimension of habit formation, where, given the role of avoidance on motivating opioid use, one could imagine a pattern of excessive avoidance but not appetitive habits in opioid addiction (see also Robbins, 2019).

In addition to habitual responses to drug-related stimuli, drug-related PIT suggests that stimuli that have acquired motivational value through Pavlovian conditioning (such as the smell of a cigarette or the sight of a syringe) can influence instrumental behaviour, either because they promote the use of a particular drug (specific PIT) or because of a general excitatory effect (general PIT). Magnitude of PIT represents the motivational properties attributed to drug-related cues, i.e. their incentive salience (the desire or "wanting" they elicit in contrast to their hedonic impact ("liking"); Berridge & Robinson, 2016; Robinson & Berridge, 1993). This has been demonstrated in regular smokers (Hogarth et al., 2010) and patients with alcohol addiction (Garbusow et al., 2016) . PIT may hold particularly important implications for cue-induced craving and relapse following abstinence, as Pavlovian associations are difficult to extinguish and can last a long time. Behavioural and neural PIT effects to alcohol-related cues in detoxified patients with alcohol addiction were able to predict relapse after one year (Sekutowicz et al., 2019; Sommer et al., 2020a). Whether a similar pattern would also emerge for patients addicted to opioids remains unknown.

After reviewing the existing evidence, I propose a bidirectional relationship between opioid addiction and aversion processing: Chronic opioid use alters processing of aversive stimuli and aversive learning due to decreases in receptor availability, receptor desensitisation or neuroplastic changes. On the other hand, maladaptive regulatory control over avoidance behaviour contributes to the development and maintenance of opioid addiction. To test this relationship, it is necessary to systematically characterise aversion processing following chronic opioid use with and without addiction. Identifying which aspects of aversion processing are altered in patients with chronic opioid use regardless of addiction and which are specific to opioid addiction will better our understanding of the relationship between aversion processing, chronic opioid use, and opioid addiction. This will help to inform patients and practitioners in their choice of medication and potentially establish novel markers to identify at-risk individuals, direct prevention efforts, and possibly identify novel targets for treating opioid addiction.

2.3 - Objective & Hypotheses

In this thesis, I seek to systematically investigate aversion processing following chronic opioid use in users with and without addiction. My objective is to shed light on the following two questions:

1. What characterises aversion processing following chronic opioid use?
2. What aspects of aversion processing are altered specifically in opioid addiction?

To answer the first question, I will compare aversion processing between patients with chronic opioid use, regardless of whether they are addicted to opioids or not, and healthy, opioid-naïve participants. To answer the second question, I will compare aversion processing between opioid-addicted patients in OMT and patients who use opioids to treat chronic pain but are not addicted to opioids. Specifically, I will use behavioural, physiological, and neuroimaging measures to investigate the following aspects of aversion processing:

- Recognition of aversive emotional expressions (Chapter 4)
- Pavlovian fear and safety learning (Chapter 5)
- Aversive instrumental learning (Chapter 6)
- Drug-related and non-drug-related Pavlovian-To-Instrumental Transfer (Chapter 7)
- Avoidance habits in contrast to appetitive habits (Chapter 8)

I expect aspects of aversion processing directly modulated by opioidergic tone to be altered following chronic use, while I expects aspects related to regulatory control over avoidance behaviour to be specifically altered in opioid addiction. Due to the direct role of the opioid system in the recognition of aversive emotions, Pavlovian fear and safety learning, and instrumental avoidance learning, I hypothesise that these processes are altered in all patients with chronic opioid use, regardless of whether they are addicted to opioids or not, compared to healthy participants. I further hypothesise that regulatory control over avoidance behaviour, demonstrated by a shift towards avoidance habits and increased aversive PIT on avoidance behaviour, is impaired only in the opioid-addicted group, with no differences between patients with chronic opioid use without addiction and healthy volunteers.

Chapter 3 - General methods

In the following, I will describe the sample of participants, methods, and procedures of the COBB ("Chronic effects of Opioid use on Brain and Behaviour") study, the results of which constitute the empirical basis for this thesis. More detailed descriptions of experimental tasks as well as additional information on data processing and statistical procedures will be provided in the respective chapters.

3.1 - Participants

In total, 86 participants were included into the final study sample, of whom 48 (56%) has current and chronic opioid use. This group consisted of 23 opioid-addicted patients in OMT (27%) and 25 non-addicted patients with opioid use for the treatment of chronic pain (29%). Additionally, 38 (44%) healthy participants without any lifetime opioid use served as a control group. One participant from the opioid-addicted group was tested but excluded from final analyses due to a childhood diagnosis of attention deficit hyperactivity disorder and one patient from the group of non-addicted patients was excluded due to regular use of neuroleptic medication. See Table 2 for demographic characteristics of the study sample.

Table 2. *Demographic characteristics of the study sample*

	Control Group	**Patient Groups**		**Group Comparison**			**Planned Contrast**	
		Non-addicted	Addicted	ANOVA/χ^2 - test			Control ≠ Patients	Non-addicted ≠ Addicted
	Mean (±SD)	Mean (±SD)	Mean (±SD)	Test statistics				
Demographics				F/χ^2	df	*p*	*p*	*p*
Sample (n)	38	25	23					
Age (years)	41.3 (11.1)	46.6 (10.0)	44.3 (10.8)	1.9	2, 83	.149	.075	.459
Gender (% female)	55%	76%	30%	10.0	2	.007	.919	.002
Handedness (% right-handed)	87%	92%	87%	1.2	4	.883	-	-
Native speaker (% German)	92%	96%	96%	0.5	2	.763	-	-
Marital status (% single/married/ divorced-widowed)	66/26/8%	48/28/24%	57/24/16%	5.1	6	.525	-	-
Body Mass Index (self-reported)	24.1 (2.7)	29.0 (7.0)	25.7 (4.6)	7.6	2, 83	<.001	.003	.021
Formal education (years)	15.3 (2.3)	12.0 (3.2)	10.1 (2.0)	34.8	2, 83	<.001	<.001	.008
Comp. education (% completed)	100%	96%	65%	20.1	2	<.001	.005	.006
Employment status (% employed)	95%	56%	65%	28.4	2	<.001	<.001	.514
Lifetime employment (years)	10.2 (10.9)	17.0 (11.2)	8.9 (8.0)	4.5	2, 83	.015	.223	.009
Disposable personal income (€ per month)	1018.4 (1006.8)	773.2 (610.5)	397.4 (586.4)	7.4	2, 83	.001	.114	<.001

All non-addicted patients reported a diagnosis of chronic pain as reason for opioid use. More specifically, patients from this group reported diagnoses of chronic back pain (44%), joint pain (24%), fibro-myalgia (24%), rheumatoid arthritis (8%), neuropathic pain (8%), or other pain syndromes (8%), with multiple selection possible. In addition, opioid-addicted and non-addicted patients self-reported the following psychiatric diagnoses: major depression disorder (43%/64%), anxiety disorders (13%/8%), alcohol use disorder (30%/0%), benzodiazepine use disorder (4%/0%), post-traumatic stress disorder (4%/0%), borderline personality disorder (8%/0%), and anorexia nervosa (0%/4%). Opioid-addicted patients used levomethadone (43%), buprenorphine (28%), methadone (12%), oral morphine (12%), or intravenous diamorphine (8%), with some patients using more than one opioid drug, for OMT. Non-addicted patients used tapentadol (48%), tilidine (24%), oxycodone (16%), hydromorphone (8%), and tramadol (4%) for the treatment of chronic pain. Additionally, opioid-addicted and non-addicted patients self-reported use of other (potentially) psychoactive medications: non-opioidergic pain medication (22%/68%), antidepressants (13%/60%), γ-aminobutyric acid (GABA) agonists/analogues (9%/44%), benzodiazepines/z-drugs (0%/16%), neuroleptics (13%/0%), promethazine (4%/8%), triptanes (4%/8%), and anti-epileptic medication (0%/4%). See Table 3 for clinical, opioid use, and other drug-use related sample characteristics.

Table 3. Clinical, opioid use, and other drug-use related sample characteristics

	Control Group	**Patient Groups**		**Group Comparison**			**Planned Contrast**	
		Non-addicted	Addicted	ANOVA/χ^2 - test			Control ≠ Patients	Non-addicted ≠ Addicted
	Mean (±SD)	Mean (±SD)	Mean (±SD)	Test statistics				
Clinical characteristics				F/χ^2	df	*p*	*p*	*p*
Acute pain (range: 1 - 10)	0.4 (0.8)	4.4 (2.5)	1.3 (2.1)	39.7	2, 83	<.001	<.001	<.001
Past-week chronic pain (range: 1 - 10)	1.7 (1.6)	6.9 (1.8)	3.6 (2.9)	49.6	2, 83	<.001	<.001	<.001
Depressive symptoms (BDI-II total score)	3.6 (3.4)	13.4 (9.5)	9.4 (8.2)	15.3	2, 83	<.001	<.001	.032
Anxiety (STAI-SF-S score)	6.5 (2.0)	7.0 (2.1)	6.9 (1.7)	2.1	2, 83	.571	.307	.817
Perceived stress (PSS-10 total score)	11.9 (5.9)	17.4 (7.1)	17.2 (5.8)	7.9	2, 83	<.001	<.001	.901
Opioid use characteristics				F/χ^2	df	*p*	*p*	*p*
Daily opioid dose* (morphine mg equivalent)	-	65.1 (51.8)	241.4 (353.2)	2.5	46	.017		
Age of first opioid use (years)	-	39.4 (9.1)	20.7 (7.7)	7.7	46	<.001		
Duration of opioid use (years)	-	7.2 (5.5)	23.6 (12.5)	6.0	46	<.001		
Prescription opioid misuse (POMI total score)	-	0.2 (0.7)	-					
Current/lifetime heroin use (%)	-	0%	13/87%					
Current/lifetime intravenous opioid use (%)	-	0%	0/57%					
Number of opioid detoxifications	-	1.5 (0.8)$^{n=2}$	4.8 (4.7) $^{n=22}$					

Number of opioid overdoses	-	-	2.7 (2.3)$^{n=9}$					
Duration of OMT (years)	-	-	14.0 (9.4)					
Duration of heroin abstinence (years)	-	-	10.3 (9.6)$^{n=20}$					
Obsessive-compulsive heroin use (OCDUS total score)	-	-	6.1 (9.3)					
Drug use characteristics				F/χ^2	df	*p*	*p*	*p*
Smoking status (% tobacco smoker)	5.3%	32.0%	87.0%	47.5	4	<.001	<.001	<.001
Age of first cigarette (years)	17.2 (4.1)$^{n=10}$	16.5 (1.9)$^{n=15}$	13.6 (5.7)$^{n=23}$	3.1	2, 45	.054	<.001	.054
Severity of tobacco use (FTND total score)	1.0 (1.4)$^{n=2}$	2.6 (1.9)$^{n=5}$	3.9 (2.4)$^{n=20}$	1.5	2, 24	.234	<.001	.790
Age of first time drunk (years)	17.5 (2.5)$^{n=33}$	17.2 (2.5)$^{n=21}$	16.5 (7.4)$^{n=23}$	0.3	2, 73	.715	<.001	.591
Past 12 month alcohol use (%)	89.5%	44.0%	65.2%	29.2	8	<.001	<.001	.210
Severity of alcohol use (AUDIT total score)	2.9 (1.7)$^{n=34}$	2.3 (1.9) $^{n=11}$	8.5 (8.3) $^{n=15}$	9.6	2, 57	<.001	<.001	<.001
Age of first cannabis use (years)	19.6 (5.0)$^{n=14}$	22.1 (9.6)$^{n=11}$	14.5 (2.4)$^{n=21}$	5.1	2, 43	.010	<.001	.005
Past 6 month cannabis use (%)	7.9%	4.0%	65.2%	43.6	8	<.001	.016	<.001
Severity of cannabis use (CUDIT-R total score)	2.7 (0.6)$^{n=3}$	5$^{n=1}$	10.0 (3.9)$^{n=15}$	10.1	1, 16	.006	-	-
Lifetime stimulant use (%)	2.6%	8.0%	95.7%	67.7	2	<.001	<.001	<.001
Lifetime benzodiazepine use (%)	2.6%	28.0%	91.3%	50.9	2	<.001	<.001	<.001
Lifetime ketamine use (%)	0%	0%	30.4%	20.9	2	<.001	.014	<.001
Lifetime psychedelic use (%)	0%	8.0%	69.6%	45.5	2	<.001	<.001	<.001
Additional risk factors								
Family history of alcohol abuse (%)	34.2%	64.0%	73.9%	10.6	2	.005	.001	.459
Family history of illicit drug abuse (%)	13.2%	24.0%	47.8%	9.1	2	.011	.019	.085
Childhood abuse (CTQ abuse subscore)	15.4 (3.1)	21.4 (10.8)	19.5 (7.9)	5.5	2, 83	.006	.002	.360

* morphine milligram equivalent dose calculated according to Nielsen et al. (2016); Abbreviations in alphabetical order: AUDIT: Alcohol Use Disorder Identification Test (Saunders et al., 1993); BDI-II: Beck Depression Inventory II (Beck et al., 1996); CTQ: Childhood Trauma Questionnaire (Bernstein et al., 2011); CUDIT-R: Cannabis Use Disorder Identification Test Revised (Adamson et al., 2010); FTND: Fagerström Test for Nicotine Dependence (Heatherton et al., 1991); OCDUS: Obsessive-Compulsive Drug Use Scale (Franken et al., 2002); OMT: opioid maintenance treatment; POMI: Prescription Opioid Misuse Inventory (Knisely et al., 2008); PSS-10: Perceived Stress Scale (S. Cohen et al., 1983);STAI-SF-S: Spielberger State Trait Anxiety Inventory – Short Form - State (Marteau & Bekker, 1992).

Inclusion and exclusion criteria

Individuals between 20 and 60 years of age and of all genders were eligible to participate in the study. The following exclusion criteria applied for all three groups: insufficient German proficiency to provide informed consent and ensure task comprehension; history of neurological or neuro-developmental disorder, including severe head trauma; severe impairments of eyesight, hearing or motor

performance; non-zero breath alcohol level or positive urine screen for stimulant drugs on the day of testing.

Healthy participants additionally were excluded if they had any personal or family history of chronic pain; personal or family history of drug addiction; any history of regular stimulant drug use (for medical or non-medical purposes); current or past harmful alcohol or cannabis use, defined as a score < 8 on the Alcohol Use Disorder Identification Test (AUDIT; Saunders et al., 1993) and a score < 6 on the revised version of the Cannabis Use Disorder Identification Test (CUDIT-R; Adamson et al., 2010); significant subclinical levels of depression or anxiety, defined as a score above cut-off levels for moderate severity (> 14 on the depression subscale, > 10 on the anxiety subscale) on the Depression-Anxiety-Stress Scale (DASS-21; Lovibond & Lovibond, 1995); positive urine screen for opioids or benzodiazepines on the day of testing.

Opioid-addicted patients had to currently be enrolled in an OMT program. In addition, patients were excluded if they had any current co-morbid non-opioid addiction; long-time regular use of stimulant drugs (e.g. cocaine, crack, amphetamine, methamphetamine, prescription stimulants); past year stimulant use; negative urine screen for their respective opioid medication on the day of testing; pronounced symptoms of opioid withdrawal at the time of testing, defined as a score > 3 on the Short Opiate Withdrawal Scale (SOWS; Gossop, 1990).

Non-addicted patients had to currently use opioid medication for the treatment of non-cancer pain for at least three months, denoting chronic use (Chou, 2015), and must not meet criteria for opioid misuse (defined as a score ≤ 2 on the Prescription Opioid Misuse Scale (POMI; Knisely et al., 2008) Additionally, patients were excluded if they had any history of drug addiction; any history of regular stimulant drug use (for medical or non-medical purposes); current or past harmful alcohol or cannabis use; negative urine screen for their respective opioid medication on the day of testing; pronounced symptoms of opioid withdrawal at the time of testing (SOWS > 3).

Recruitment strategy

Different recruitment strategies were applied depending on the study group: Opioid-addicted patients were recruited through posters and leaflets distributed to addiction treatment facilities in and around Hamburg, as well as personal referral by their attending physicians, nurses or social workers and word-of-mouth. Non-addicted patients were also recruited through posters and leaflets disseminated to specialised pain treatment practices, hospital units, rehabilitation services and self-help groups, or by physicians' referral. We advertised study participation to healthy participants using an online platform (Stellenwerk, *www.stellenwerk.de*) for research volunteers, leaflets distributed in public and

residential areas as well as word-of-mouth. Interested candidates for this group enlisted on an online survey platform (LimeSurvey Community Edition Version 5.5.0, *www.limesurvey.org*), where they entered demographic information and also filled out the DASS-21 to screen for sub-clinical levels of depressive symptoms and anxiety, as well as the AUDIT and CUDIT-R to screen for current harmful alcohol or cannabis use. See Figure 1 for a flow chart depicting the recruitment and screening process.

Screening procedure

Prior to inclusion into the study, we conducted telephone screenings with all candidates to evaluate their eligibility to participate. These screenings queried demographic information, (comorbid) psychiatric diagnoses and medical history, current/past medication and drug use as well as MRI eligibility. Candidates for the two patient groups (who did not enlist online in order to remove as many barriers to study participation as possible) completed AUDIT and CUDIT-R during this phone screening. Candidates for the non-addicted patients also completed the POMI to assess lifetime misuse of opioid medication. Additionally, current smokers completed the Fagerström Test for Nicotine Dependence (FTND; Heatherton et al., 1991).

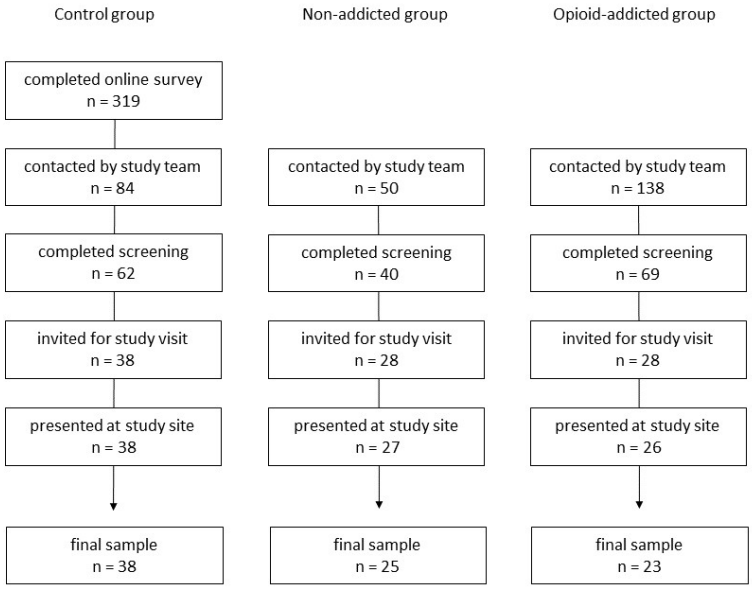


Figure 1. Recruitment flow chart for all three groups of participants.

3.2 - Study procedure

The study was conducted in adherence to the ethical principles for medical research in human participants outlined in the Declaration of Helsinki. Data collection was carried out between December 2021 and September 2024 at the University Medical Centre Hamburg-Eppendorf's Department of Systems Neuroscience in Hamburg, Germany. Following successful screening, study participants

selected their preferred appointment for the study visits. Whereas a minority of participants (7%; 5 non-addicted patients, 1 opioid-addicted patient) completed the study in one session due to long travel time, most completed the study in two parts. Every participant received comprehensive written information prior to the study visit and patients were instructed to take their opioid medication as they usually would when participating in the study.

After arriving for their (first) study visit, participants were verbally informed about the study contents and objectives as well as confidentiality/data protection policies by the experimenter and provided written informed consent. Participants were then tested for study-day exclusion criteria by measuring breath alcohol level using a hand-held breathalyser (ACE DA-5000; ACE, Freilassing, Germany) and performing urine drug screens for the following substances: morphine/opiates, methadone, buprenorphine, cocaine, amphetamine, and benzodiazepines (Drug Test; nal von minden, Moers, Germany). If non-addicted patients reported use of an opioid that is not a morphine derivate – and would therefore not be detected by the morphine/opioid screen test – that medication was additionally tested for. In total, three participants were excluded on the day of testing: one patient from the opioid-addicted group had a breath alcohol level > 0; one patient from the opioid-addicted group had a positive urine screen for cocaine; one patient from the non-addicted group had discontinued her prescribed opioid about one week prior to testing. Participants from the patient groups then provided time and dosage of last opioid intake and completed the SOWS in an interview format to screen for opioid withdrawal symptoms. After confirming inclusion eligibility, participants performed a fixed order of behavioural tasks (see Table 4) and questionnaires presented on two laptops with touchscreen-capable screens (DELL Inspiron 7506 2n1), wearing connected headphones if required by the task. This concluded the first study visit (approximate duration: 3.5 hours), and participants received the first portion of their reimbursement. In what was the second study visit for most (although some completed this part immediately after the interviews), participants were instructed about the MRI procedure and declared suitable for MRI by a specialised physician. Inside the scanner room, a respiratory belt and finger pulse oximeter were applied to record physiological data during the MRI session. Two MRI compatible Ag/AgCl electrodes recorded electro-dermal activity from the left palm (see below). To exert electrical stimulation required for the two tasks carried out in the MRI setting, a Wasp stimulation electrode wrist (Digitimer Ltd., Letchworth Garden City, UK) was applied on the inside of each wrist (cable grounded through a radio frequency filter panel). For all participants, shocks were applied for 200 ms at a maximum voltage of 400 V, while the amperage was individually determined prior to the start of the task using a shock work up procedure. During this, an initially very low electrical pulse (0.1 mA) was repeatedly delivered and continuously increased in steps of 0.5 mA until the participant experienced the shock intensity as very unpleasant but not yet painful. MRI recording commenced with closed-eyes resting state MRI before initiating the first behavioural

task (outcome devaluation task; see Chapter 8). Completion of the task was followed by the anatomical MRI scan, and a second task (threat reversal task; see Chapter 5; see below for specific scanning parameters). Altogether, participants remained inside the scanner for approximately 90 minutes. Participants received the remainder of their monetary reimbursement. In total, healthy participants received €50 and - due to longer time of testing - patients received €70 €, plus travel expenses. Afterwards, participants were thanked for their participation and saw to their way of transportation if necessary.

3.3 - Behavioural tasks

For an overview over the behavioural tasks used in this thesis to investigate different aspects of aversion processing see Table 4. Detailed task descriptions and information on data collection, processing, and analysis can be found in the respective chapters. Two additional tasks (Stop Signal Task and the Monetary Choice Questionnaire) are not included in this thesis.

Table 4. Overview over behavioural tasks with corresponding outcome measures

Construct	Task Name	Behavioural outcome measures	Additional outcome measures	Chapt.
Facial Emotion Recognition	Emotional intensity morphing task*	Point of emotional intensity required to recognise expressions of five basic emotions (anger, disgust, fear, happiness, sadness) in other's faces.	-	4
Fear Reversal Learning	Threat Reversal Task**	No behavioural measures	fMRI BOLD and skin conductance responses to CS+ and CS- before and after reversal of threat and safety cues.	5
Instrumental Avoidance Learning	Instrumental avoidance learning task using aversively conditioned stimuli	Accuracy and latency of avoidance behaviour	Skin conductance responses to conditioned cues.	6
Pavlovian-To-Instrumental Transfer	PIT paradigm with drug-related or neutral conditioned stimuli	Effects of conditioned cues on accuracy and latency of instrumental behaviour (PIT effect).	Differences between neutrally and drug-related conditioned cues (context effect).	7
Habit Learning	Overtraining of instrumental avoidance responses and habits tested by outcome devaluation task***	Accuracy and latency of avoidance responses during learning and in extinction.	fMRI BOLD and skin conductance responses to conditioned cues before and after outcome devaluation.	8
	Appetitive contingency degradation task****	Response rates under different instrumental contingencies (non-degraded, partially degraded, fully degraded).	-	
	Avoidance contingency degradation task	Response rates under different instrumental contingencies (non-degraded, partially degraded, fully degraded).	-	
Behavioural inhibition	Stop Signal Task◊	Stop-Signal-Reaction-Time, measuring response-inhibition latency.	-	-
Choice impulsivity	Monetary Choice Questionnaire◊◊	Discounting rate describing choice behaviour when choosing between small-immediate and large-delayed rewards.	-	-

* see Bland et al. (2016); ** see Apergis-Schoute et al. (2017); *** see Gillan et al. (2015); **** see Ersche et al. (2021);
◊ task data not included in this thesis, see Verbruggen et al. (2019); ◊◊ task data not included in this thesis; see Kirby et al. (1999)
BOLD: blood oxygen level dependent

3.4 - Physiological recordings and pre-processing

Electro-dermal activity (EDA) can be used for quantifying physiological arousal. Increases in arousal lead to sudomotor responses (i.e. sympathetic innervation of the sweat glands) that change the conductance of electrical activity on the skin (Critchley, 2002). The EDA signal represents the recording of skin conductance, which consist of two components: the tonic (i.e. slowly changing) skin conductance level (SCL) and the phasic (rapidly changing) skin conductance responses (SCR). SCR is of interest in cognitive and clinical neuroscience because it provides an index for the physiological responses to presented stimuli.

In this study, participants' EDA was acquired both within and outside the MRI setting using two Ag/AgCL electrodes placed on the participants' non-dominant palm (over the hypothenar muscle). For the task performed outside the MRI environment (aversive learning task, see Chapter 6), the EDA signal was amplified and converted to digital using a BIOPAC MP100 acquisition system (1000 Hz sampling rate) and recorded using AcqKnowledge 3.9.1 (both BIOPAC Systems Inc., Goleta, CA, USA). Prior to the start of the task, 90 seconds of resting-state activity was recorded in order to reliably determine baseline skin conductance and to avoid artefacts in early-trial recordings. For the tasks performed in the MRI environment (outcome devaluation task and threat reversal task, see Chapters 5 and 8, resp.), the transduction cable connecting the electrodes to the recording device was grounded through a radio frequency filter panel; the analogue signal (again sampled at 1000 Hz) was then amplified (BIOPAC MP150; BIOPAC Systems Inc., Goleta, CA, USA) and converted into digital using a CED Micro1401-3 data acquisition unit (Cambridge Electronic Design Ltd., Cambridge, UK). Accompanying Spike2 software (version 6.17) was used for data acquisition. Baseline skin conductance was recorded during the preceding MRI sequences.

Pre-processing of EDA data and computation of SCR was carried out using the Ledalab toolbox (version 3.4.9; Benedek & Kaernbach, 2010) implemented in MATLAB (version R2020b; The MathWorks Inc., Natick, MA, USA). In a first step, EDA data were down-sampled to 100 Hz and trimmed to the task duration. Then, all EDA recordings were visually inspected. Participants with bad signal-to-noise ratio where no change in EDA in response to the presentation of the US (aversive images/sounds and electrical stimulation, respectively) was apparent were excluded from the analysis. To determine SCL, reflective of tonic changes in EDA signal, EDA time courses of included participants were subject to a first-order low-pass Butterworth filter (cut-off at 5 Hz). Ledalab uses a mathematical approach called Continuous Decomposition Analysis to parse out tonic (SCL) and phasic (SCR) changes in EDA (Benedek & Kaernbach, 2010). SCL is subtracted from the EDA signal, which is then temporally smoothed using a Gaussian kernel (mean: 0 ms ± 200). Finally, SCR to different stimuli were calculated as the signed event-related signal change in pre-defined response windows following stimulus onset (see Chapters

5 and 8) according to conventional peak-detection windows for fear conditioning tasks (Kuhn et al., 2022). Minimum SCR amplitude was set to 0.01 µS. SCR for each participant and each trial were then extracted for further analysis ("ERA.CDA.SCR" value), z-transformed for each participant separately to compare responses to different stimuli regardless of inter-individual differences, and log-transformed to normalise the data ($log(z_{SCR}+1)$).

3.5 - MRI acquisition and pre-processing

Magnetic resonance imaging (MRI) is a non-invasive method to create detailed images of internal structures such as the brain (Glover, 2011). In functional MRI (fMRI), neural activity can be temporally and regionally estimated by leveraging changes in brain oxygen metabolism in response to external event (blood-oxygen level dependent imaging; BOLD; Logothetis, 2003). In the present study, functional and anatomical imaging was performed on a PRISMA 3T MRI scanner (Siemens, Erlangen, Germany), using a 64-channel head coil. An anatomical image was obtained for analysis of structural differences between the groups (not included in this thesis) as well as for co-registration and subsequent normalisation of functional data. Using a magnetisation prepared rapid gradient echo (MPRAGE) sequence, a T1-weighted image was acquired with 176 slices recorded, a field-of-view (FoV) of 240 mm x 256 mm x 176 mm, a voxel-size of 1.0 x 1.0 x 1.0 mm, a recording time (TR) of 2300 ms, and an echo time (TE) of 2.98 ms. A second anatomical image as acquired using a fluid-attenuated inversion recovery sequence with 27 slices recorded, FoV of 224 mm x 182 mm x 134 mm, voxel-size of 0.7 x 0.7 x 4.0 mm, TR of 7840 ms, and TE of 96 ms (not included in this thesis). A total of four runs of functional images were recorded: one run of resting-state data (not included in this thesis), and three runs for the two behavioural tasks. For the first task (avoidance habit learning task, see Chapter 8), two runs were recorded, one for the four training blocks and a one for the test block. Both runs used identical scanning parameter: 1270/296 volumes (training/test) of T2*-weighted images were recorded using a multi-slice echo-planar imaging (EPI) sequences, each volume consisting of 48 slices, FoV of 224 mm x 224 mm x 145 mm, a flip angle of 66°, voxel-size of 2.0 x 2.0 x 2.0 mm, TR of 1420 ms, TE of 25 ms, multiband factor of 2, and in-plane acceleration factor of 2. For the second task (fear reversal paradigm, see Chapter 5), 785 volumes of T2*-weighted images were recorded using a multi-slice EPI sequences, each volume consisting of 58 slices, FoV of 224 mm x 224 mm x 145 mm, a flip angle of 70°, voxel-size of 2.0 x 2.0 x 2.0 mm, TR of 1710 ms, TE of 25 ms, multiband factor of 2, and in-plane acceleration factor of 2. For the resting-state data, 240 volumes of T2*-weighted images were recorded using a multi-slice EPI sequence, each volume consisting of 64 slices, FoV of 224 mm x 224 mm x 145 mm, a flip angle of 74°, voxel-size of 2.0 x 2.0 x 2.0 mm, TR of 1880 ms, TE of 25 ms, multiband factor of 2, and in-plane acceleration factor of 2. To correct for possible field inhomogeneity, a 3D-

gradient echo field map was acquired, with 48 slices, FoV of 224 mm x 224 mm x 145 mm, voxel-size of 3.0 x 3.0 x 3.0 mm, TR of 483 ms, and TE_1 of 4.62 ms and TE_2 of 7.08 ms.

Pre-processing of the task-related fMRI data was carried out using SPM 12 (version 12; Wellcome Department of Cognitive Neurology, London, UK) implemented in MATLAB (version R2020b; The MathWorks Inc., Natick, MA, USA). All images were imported, converted to NIFTI format, and organised as 4D-files according to the Brain Imaging Data Structure (BIDS; Gorgolewski et al., 2016). EPI volumes were then temporally corrected using SPM's slice-time correction, realigned to the mean EPI using a two-pass procedure, segmented into grey matter, white matter, and cerebro-spinal fluid (CSF), and registered to the anatomical T1-weighted image using a nonlinear co-registration approach implemented in the CAT12 toolbox for SPM (Gaser et al., 2024). Mean EPI and T1-weighted image were then skull-stripped to only contain grey matter, white matter, and CSF voxels and subsequently warped into a common space using DARTEL-based transformation fields (Ashburner, 2007) and templates including in CAT12. Successful normalisation of the mean EPI into template space was visually inspected and verified for all participants. First- and second-level analysis of fMRI data was task-specific and is described in the corresponding task chapters.

Associations of task measures and opioid use characteristics

To further explore the relationship between aversion processing and opioid use, a subset of variables related to patients' opioid use were correlated with the primary task outcome measures: daily morphine milligram equivalent dose, age of first opioid use, duration of opioid use, and, for opioid-addicted patients only, severity of obsessive-compulsive drug use (OCDUS total score).

Calculation of the required sample size

Required sample size was optimised for a detecting behavioural group difference in one of the two fMRI tasks included in this thesis. A previous study that tested differences between patients and healthy participants in the same task served as reference (Gillan et al., 2015). The observed group difference of the primary outcome measure in that study was 15.2 (± 23.7). Given $\alpha = .05$ and $\beta = .80$, minimum sample was estimated to be n = 39. Therefore, 40 participants per group were targeted. However, due to time constraints and difficulties in patient recruitment, it was decided to stop ahead of reaching that goal.

3.6 - Statistical analyses

Data were prepared using MATLAB (version R2020b; The MathWorks Inc., Natick, MA, USA) and analysed using SPSS (version 29.0.2; IBM Corp., Armonk, NY, USA). Continuous demographic, clinical and drug-use-related variables were compared between the groups using one-way ANOVA, with planned contrast for group comparisons testing either opioid-using patients against healthy participants (Contrast 1: 1 -.5 -.5) or testing the two patient groups against each other (Contrast 2: 0 1 -1). In case of unequal variances between the groups (indicated by a significant result in Levene's test), results were reported following Brown-Forsythe correction, which accounts for group differences in variance. On variables where only the two patient groups were compared, I used independent samples *t*-tests. Group differences in nominal variables were analysed using χ^2-tests, with post-hoc group comparisons between healthy participants and patients or between opioid-addicted and non-addicted patients reported as Bonferroni-corrected *z*-tests. Depending on the specific hypothesis, task-related data was tested for either 1) group differences between healthy participants and patients, for which opioid-addicted and non-addicted patients were collapsed or 2) group differences between opioid-using patients with addiction and opioid-using patients without addiction. For this, groups were compared using one-way ANOVA (see above) or repeated measures ANOVA. If suitable, covariates were added. If, in the latter case, the assumption of sphericity was violated, degrees of freedom and *p*-values were reported following Greenhouse-Geisser correction, accounting for uneven variance on different factor levels. Depending on the hypothesis, planned contrasts were applied on either main or interaction effects of interest. For post-hoc comparisons, the Tukey test was used on one-way ANOVA and Bonferroni-Holm correction was applied when repeated measures ANOVA were used. Cohen's *d* and partial η^2 were used to indicate effect sizes, with $d = |0.20|/\eta^2 = .01$ indicating a small effect size, $d = |0.50|/\eta^2 = .06$ indicating medium effect size and $d = |0.80|/\eta^2 = .14$ indicating a large effect size (J. Cohen, 2013; J. Miles & Shevlin, 2014). Relationships between task outcome measures and opioid use characteristics as well as known risk factors for developing opioid addiction were tested using Pearson's correlation for both patient groups independently, with $r = |.30|$ indicating a small correlation, $r = |.50|$ indicating a medium correlation, and $r = |.70|$ indicating a large correlation. For variables with extremely skewed distributions, such as physiological or neuroimaging outcome measures, Spearman correlation was used, which is rank-based and therefore less affected by extreme values. Spearman's ρ can be interpreted similarly to Pearson's *r*. All statistical tests were two-tailed with $\alpha = .05$.

Chapter 4 - Recognition of aversive emotional expressions

4.1 - Introduction

Successful social interaction requires the ability to identify the current emotional state of another person and adapt the own behaviour accordingly, e.g. by taking a step back when recognising anger in another person or changing to a more soothing tone of voice when recognising fear. Facial expressions are among the most informative sources for inferring other's emotions (Barrett et al., 2019) and deficits in recognising emotions from faces, especially aversive emotions like anger and fear, have been linked to reduced empathy (Besel & Yuille, 2010), impairments in perspective-taking (Trubanova et al., 2015), and psychopathic personality traits (Dawel et al., 2012). As such, recognition of aversive emotions from facial expressions can be considered a fundamental building block for social intelligence and the formation of functional interpersonal relationships.

Forming and maintaining social relationships is particularly crucial for patients with opioid addiction. Social support and integration into social networks are associated with higher probability of abstinence and improved outcomes of addiction treatment (Cavaiola et al., 2015; Kumar et al., 2021); social isolation on the other hand has been linked to both the development and exacerbation of opioid addiction (N. C. Christie, 2021). Impairments in social cognition (McDonald et al., 2013) and empathic concern (Tomei et al., 2017) in opioid addiction can put a strain on relationships and could therefore potentially contribute to the maintenance of addiction, hindering social support and fostering isolation.

It has been proposed that an impaired ability for emotion recognition could underlie social difficulties in opioid addiction. An early study found that patients with opioid addiction (either in OMT or recently detoxified) displayed lower emotion recognition accuracy compared to healthy controls when asked to select the emotion expressed in photographs of faces (Kornreich et al., 2003). Results indicated a general rather than emotion-specific deficit, with impaired recognition across all emotions. A conceptual replication did not find reduced accuracy in opioid-addicted patients but patients in OMT were slower to identify emotions compared to controls and abstinent patients (Martin et al., 2006). More recent studies with similar methodology reported reduced recognition accuracy and recognition speed in opioid addicted patients in OMT, again without differences between emotions (Craparo et al., 2016; Terrett et al., 2020). Craparo et al. (2016) additionally found a relationship between facial emotion recognition and self-reported alexithymia, i.e. the ability to identify, describe, and distinguish between feelings (Sifneos, 1973), which has been repeatedly linked to drug addiction (Honkalampi et al., 2022), and has been proposed as a risk factor for developing opioid addiction (Henschel et al., 2022; Oyefeso et al., 2008). Applying an adapted task design, one study showed opioid-addicted patients

video clips of actors portraying different emotions and asked them to identify the emotion in question (McDonald et al., 2013). Similar to the other studies, emotion recognition was reduced across emotions in patients in OMT; patients who were abstinent for at least three months, however, did not differ from control participants. Further, the study found that emotion recognition was mediated by overall cognitive function, which the authors suggested to be acutely impaired by the methadone used for OMT. This does not, however, account for earlier findings of reduced emotion recognition in recently detoxified patients and instead implies recognition deficits as a consequence of chronic opioid use that recover following prolonged abstinence.

Chronic opioid use is known to lead to structural and functional adaptations of opioid receptors (M. J. Christie, 2008), altering neuro-cognitive mechanisms reliant on opioidergic modulation in the process (Al-Hasani & Bruchas, 2011). Evidence from experimental studies in healthy volunteers implicates the opioid system in recognising aversive emotions from facial expressions: When opioidergic activity was reduced by naltrexone-induced receptor blockage, participants required higher intensities of emotional expressions to identify fearful and sad faces (Wardle et al., 2016). Interestingly, stimulating the opioid system using opioid agonists also reduced the ability to recognise aversive emotions: sensitivity for angry facial expressions was attenuated following morphine (Løseth et al., 2018) and identification of fearful expressions was impaired following buprenorphine (Ipser et al., 2013). This goes to show that a balanced functioning of the opioid system is critical for efficient emotion recognition. Mechanistically, manipulating opioidergic tone has been demonstrated to alter spontaneous gaze patterns towards the eye and nose/mouth/jaw regions (Chelnokova et al., 2016), with either too much or too little exploration of these regions potentially explaining inefficient emotion recognition.

Taken together, these findings suggest that altered function of the opioid system following chronic use could explain the findings of impaired emotion recognition in opioid-addicted patients. However, the aforementioned study designs do not ultimately allow for an interpretation about the directionality of this effect. While it is likely that neural adaptations related to chronic opioid use lead to impaired emotion recognition, it could also possible that impaired emotion recognition predates the onset of addiction. Testing both addicted and non-addicted patients with chronic opioid use could help differentiate the two alternatives: Similar impairments in patients irrespective of addiction would indicate emotion recognition deficits related to chronic use, while a diverging pattern between the two groups would point to an addiction-specific effect, which might potentially serve as a risk factor for developing opioid addiction.

Aims & Hypotheses

In this chapter, I aim to clarify the relationship between chronic opioid use and emotion recognition by comparing addicted and non-addicted patients with chronic opioid use to healthy participants. I hypothesise that the recognition of emotions from facial expressions is impaired in both patient groups, indicated by a higher intensity of emotional expression required for recognition. I further hypothesise that the two patient groups will not differ from each other in terms of emotion recognition.

4.2 - Methods

Sample

All 86 participants included in the study completed the facial emotion recognition task and were subsequently part of the analyses reported in this chapter.

Task description

The Emotional Intensity Morphing Task (EIMT), included in the emotion processing subsection of the EMOTICOM test battery (Bland et al., 2016), can be used to detect the intensity required to recognise emotions in facial expressions. In contrast to other tasks assessing the processing of emotional expressions in faces, this task does not test the identification of emotions – the emotion in question is announced to the participants before a trials starts. Rather, emotions are displayed in ascending or descending intensities with participants indicating their point of detection, testing how sensitive they are to recognising different emotions. The task has been previously demonstrated to detect emotion-specific alterations in emotion recognition intensity in a sample of patients with cocaine addiction (Bland & Ersche, 2020).

In this task, participants were instructed to observe the picture of a face that either increases (forward condition) or decreases (backward condition) in emotional intensity (see Figure 2). In the forward condition, a neutral expression was replaced with a series of 13 morphed images increasing in intensity until a high emotional expression was reached, resulting in a 15-step scale. Each image was displayed for 500 ms. Participants were asked to press the space bar once they detected the presence of an emotion. Importantly, the emotion in question was announced to the participants before every trial (e.g. “Press SPACE once you recognise ANGER”). In the backward condition the images were presented in reverse order, i.e. from high emotional intensity to a neutral facial expression, and participants were asked to press the space bar once they no longer detected the presence of an emotional expression (e.g. “Press SPACE once you no longer recognise ANGER”). The task includes five basic emotions (anger,

disgust, fear, happiness, and sadness), as well as a sixth (surprise) that was used for two practice trials prior to the start of the task (one for the forward and one for the backward condition). There were two different sets of pictures: One set was showing a female, the other one a male face. Each of the five emotions was presented twice per condition and gender, resulting in a total of 40 trials per participant. The intensity at which participants detected (forward) or no longer detected (backward) each emotion was used as the primary outcome measure, with a lower value meaning that participants recognised the emotion at a lower intensity. If participants did not press the space bar at all, intensity was set to 15 (forward) or 0 (backward), respectively.

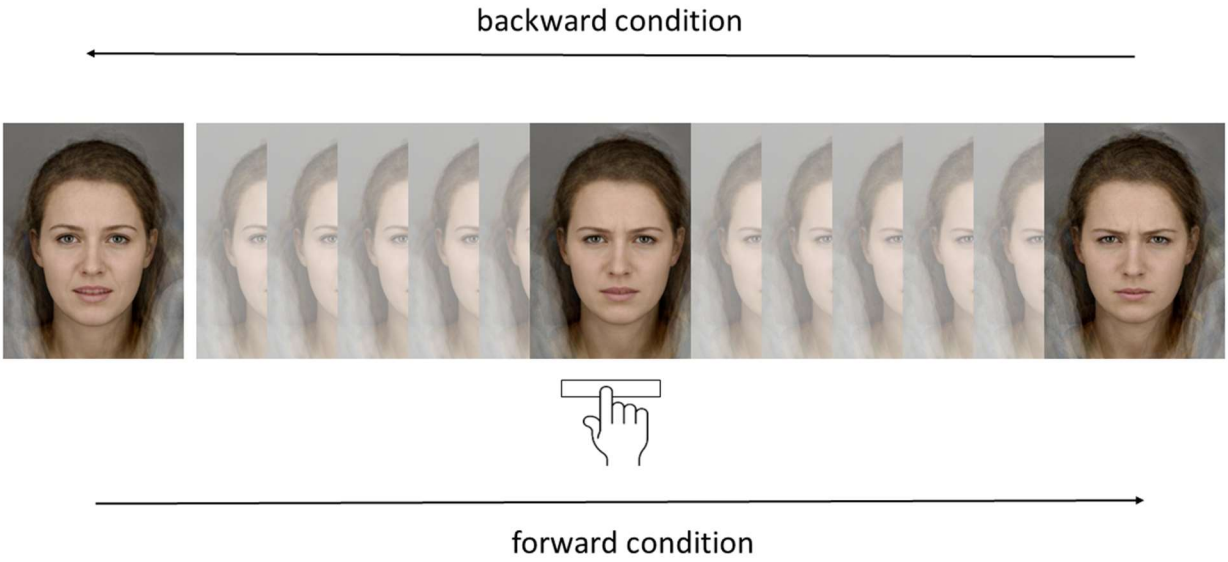


Figure 2. Schematic representation of the Emotional Intensity Morphing Task

Data analysis

Intensity values were averaged within trial types. Group differences in overall emotion recognition were tested using a 2x2x3 repeated measures ANOVA, with intensity as dependent variable, (the face's) gender (female/male), and condition (forward/backward) as within-subject factors, and group (control/non-addicted/addicted) as between-subject factor. For group comparisons, planned contrasts were a priori specified: Contrast 1 collapsed the opioid-using groups and compared them to the healthy participants (contrast weights on the main effect of group: 1 -.5 -.5); Contrast 2 compared the non-addicted to the addicted opioid-using group (contrast weights on the main effect of group: 0 1 -1). To further test for group differences in emotion-specific recognition, a multivariate 2x2x3 ANOVA was conducted, with intensities for each emotion as separate dependent variable, gender and condition as within-subject factors, and group as between-subject factor. The same contrasts were applied for planned comparisons. As behaviour in the Emotional Morphing Intensity task is not impacted by

differences in gender or years of education (Bland et al., 2016), these variables were not included as covariates into the analysis, even though the groups differed in those characteristics.

4.3 - Results

Repeated measures ANOVA revealed significant main effects for face gender ($F_{1,83}$ = 30.0, p < .001, η_p^2 =.27), and condition ($F_{1,83}$ = 193.6, p < .001, η_p^2 = .7) as well as a significant gender-by-condition interaction ($F_{1,83}$ = 30.7, p < .001, η_p^2 = .27), indicating a lower intensity required to recognise female compared to male faces in the backward direction. Analysis revealed no main effect of group ($F_{2,83}$ = 0.7, p = .498) or any significant interactions including group (all p > .5). Planned contrasts showed no differences in overall emotion recognition between the healthy participants and the opioid-using patients (t_{83} = 0.2, p = .405) or between the two opioid-using patient groups (t_{83} = -0.26, p = .417; see Figure 3).

For the emotion-specific multivariate ANOVA, planned contrasts identified no significant differences between healthy participants and patients for any emotion (anger: t_{83} = 0.18, p = .557, disgust: t_{83}= 0.44, p = .181, fear: t_{83}= -0.02, p = .939, happiness: t_{83} = 0.23, p = .467, sadness: t_{83}= 0.17, p = .618). Planned comparisons did, however, reveal a significant group difference between opioid-addicted and non-opioid-addicted patients in the intensity required to recognise fearful facial expressions (t_{83} = -0.94, p = .016, d = 0.73), indicating a higher intensity required for the opioid-addicted patients, while intensities required for recognising anger (t_{83} = -0.64, p = .117), disgust (t_{83} = 0.11, p = .802), happiness (t_{83} = 0.36, p = .387), and sadness (t_{83} = -0.18, p = .692) did not differ between the two groups (see Figure 3).

Emotion recognition measures were not significantly correlated to any opioid use characteristics for either patient group.

4.4 - Discussion

The aim of this chapter was to clarify the relationship between facial emotion recognition and chronic opioid use. Patients and healthy participants completed a task in which they indicated their point of detection for five basic emotions (anger, disgust, fear, happiness, sadness) with a button press. In a first planned comparison, patients with chronic opioid use were compared to healthy, opioid-naïve

participants and, in a second planned comparison, opioid-addicted patients were compared to non-addicted patients with chronic opioid use.

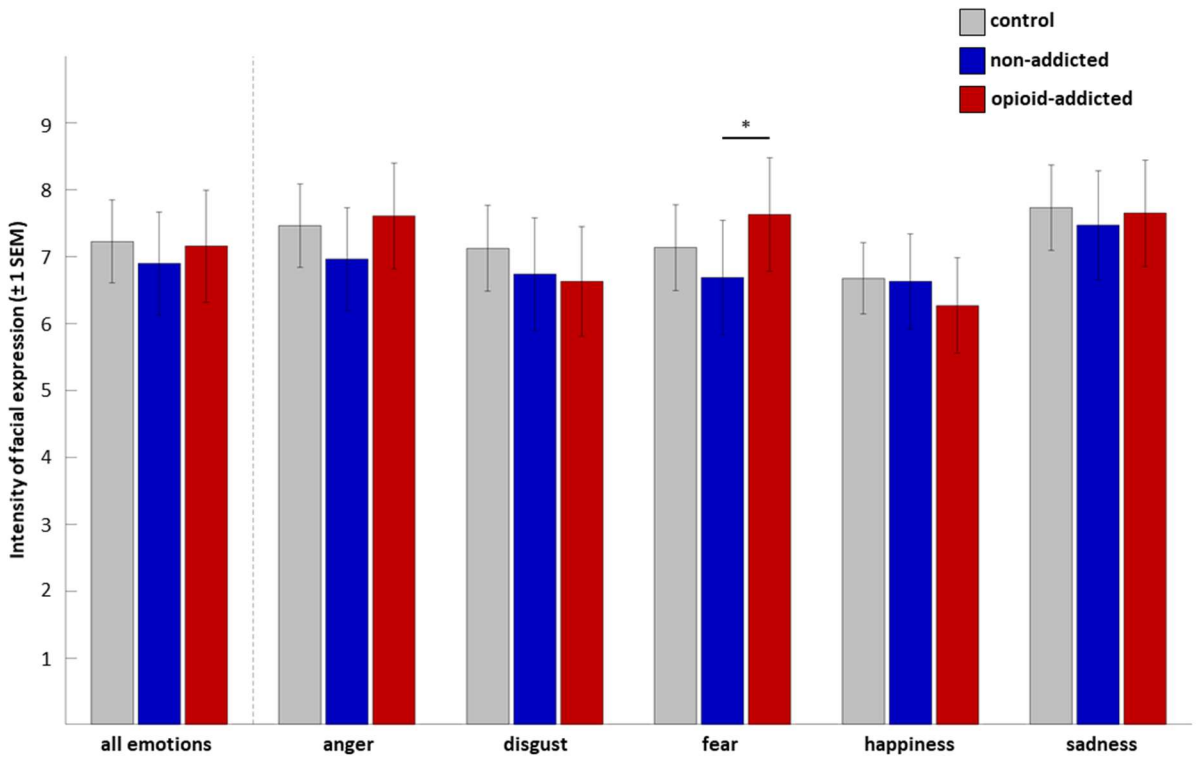


Figure 3. Overall and emotion-specific intensities required for emotion recognition.

In contrast to my first hypothesis, patients with chronic opioid use displayed no deficits in emotion recognition, marked by a point of detection comparable to the control group. This finding was unexpected due to the role opioids had been shown to play in recognising facial expressions (e.g. Wardle et al., 2016). The reduced opioidergic tone related to chronic opioid use (Kling et al., 2000) was expected to disrupt facial emotion recognition by altering spontaneous gaze patterns (Chelnokova et al., 2016). It is possible that the acute effects of the opioids currently used for either OMT or pain treatment compensate for potentially long-term adaptations caused by chronic use. A study probing the processing of fearful faces in opioid-addicted patients found hyper-activation of the amygdala, a brain structure orchestrating the inadvertent attentional shift towards the eye-region facilitating correct identification of negative emotions, when testing patients prior to acute opioid administration (Gamer & Büchel, 2009; Schmidt et al., 2014). Following heroin injection, however, amygdala responses were comparable to healthy participants who had received saline injections (Schmidt et al., 2014), which suggests that in patients with chronic use, acute opioids might "normalise" neural processes involved in facial emotion recognition. This is supported by another study that reported lower emotional empathy scores in patients in OMT, but only if their last opioid dose had been taken more than 12 hours ago and was likely not active anymore (Carlyle et al., 2020). In patients who had taken their OMT medication on the day of testing, empathy scores did not differ from control

participants. Emotional empathy relies on facial expression recognition (Besel & Yuille, 2010), suggesting a similar mechanism could be at play here. In the present study, patients were eligible to participate only following positive urine screen tests to opioids, yet showed no symptoms of opioid withdrawal, indicating they were currently subjected to effects of their prescribed opioid drug. Therefore, given the interplay between acute opioid effects and tolerance in long-term users, a balanced opioidergic tone results in intact emotion recognition in patients.

The results in this chapter contrast previous studies investigating emotion recognition from facial expression in opioid addiction, in most of which patients were also on OMT. Several studies found impairments in the identification of emotional expressions in patients, either in accuracy (Craparo et al., 2016; Kornreich et al., 2003; McDonald et al., 2013; Terrett et al., 2020) or in recognition speed (Craparo et al., 2016; Martin et al., 2006). Data from our study do not support such an overall emotion recognition deficit. This could be explained by different methods for measuring emotion recognition. Most of the studies above used some variation of the same task design, where participants were presented with pictures or videos of faces displaying a certain emotion and were subsequently asked select the correct emotion from a set of options. The EIMT used in the present study diverges from this approach and instead asks participants to press a button once they recognise (or no longer recognise) a previously declared emotion (Bland et al., 2016). As such, EIMT assesses the point of emotional intensity required to recognise a facial emotion, which is a measure of sensitivity towards detecting emotion-related facial movements. The identification of emotions in facial expressions measured by previous studies incorporates additional processes (Spunt & Adolphs, 2019), which require, among others, the retrieval of information about emotion categories. EIMT therefore isolates emotion recognition from more cognitive aspects of emotion recognition and hence provides a more sensitive measure for affective aspects of aversion processing.

In line with my second hypothesis, opioid-addicted patients did not differ from non-addicted patients in overall emotion recognition. They did, however, require a higher intensity of emotional expression to recognise fear, while recognition of other basic emotions did not differ between the two patient groups. Emotional intensity required for fear recognition was not correlated with the current opioid dose or duration of opioid use, making it unlikely that the difference between the groups is caused by higher exposure to opioids in the addicted group. Impaired fear recognition has previously been found in patients addicted to alcohol (Trick et al., 2014), methamphetamine (Kim et al., 2011), and cocaine (Bland & Ersche, 2020; Ersche et al., 2015; Morgan & Marshall, 2013), which implies a role for fear recognition deficits in drug addiction in general. Fear recognition, a key social aspect of aversion processing, could potentially be disrupted as part of the neural adaptations related to the addiction cycle, where escalating drug use results in changes to brain and behaviour. Especially the

withdrawal/negative affect stage of the model is linked to changes in aversion processing-related neurocircuitry, including the aforementioned amygdala (Koob & Volkow, 2016), suggesting that impaired fear recognition could co-develop with and potentially further exacerbate drug addiction, as it is linked to impaired socio-cognitive functioning (Corden et al., 2006) and thus may hinder the formation and maintenance of supporting relationships. At the same time, it is also possible that a decreased sensitivity for fearful faces predates the onset of drug addiction and thus serves as a risk factor for its development. Patients with opioid misuse and pronounced alexithymia, closely associated with impairments in facial emotion recognition (Craparo et al., 2016), have reported increased opioid use in distressing situations (Henschel et al., 2022). Difficulties in understanding own and others' feelings can be highly distressing and repeatedly avoiding this negative affect due to the ability of opioids to reduce social stress (Bershad et al., 2015) might contribute to the development of opioid addiction. Ultimately, from the present data, it is not possible to determine whether deficits in fear recognition in patients addicted to opioids result from addiction-related neuroadaptations or predate its development and can thus be considered as risk factor.

To summarise, patients with chronic opioid use did not differ from healthy participants in overall or specific emotion recognition from facial expressions. However, opioid-addicted patients required a higher intensity of emotional expressions to recognise fear compared to non-addicted patients. This suggests that impairments in fear recognition, a crucial sub-process of aversion processing, are specifically related to opioid-addiction and not a necessary consequence of chronic opioid use.

Chapter 5 - Pavlovian fear and safety learning

5.1 - Introduction

To safely navigate our daily surroundings, learning from experience as to what poses a threat and what conveys safe is key. From predatory animals to busy intersections, this holds true nowadays as it did throughout human history and across species: Learning what to fear and what not to fear ensures survival. Crucial mechanisms for this are Pavlovian fear conditioning, where previously innocuous stimuli are learned to elicit fear responses after being followed by a threat, and fear reversal, where learned fear and safety-related knowledge is updated and flexibly adapted to changing circumstances. Learning and updating of fear are critical aspects of aversion processing and as such are modulated by the opioid system; it remains, however, unknown how chronic opioid use affects the functioning of these mechanisms. Additionally, fear conditioning and impairments in flexibly updating fear and safety have been proposed to contribute to the chronic nature of opioid addiction by facilitating conditioned withdrawal and subsequent relapse (Pantazis et al., 2021).

In this chapter, I will first describe the mechanisms of fear conditioning and reversal, introduce their underlying neural networks, and the respective role of the opioid system. I will then theoretically link the two mechanisms to opioid addiction before using physiological and neuroimaging data to characterise fear conditioning and reversal following chronic opioid use with and without addiction.

Fear conditioning

In the terminology of experimental psychology, fear conditioning occurs when an aversive event (unconditioned stimulus (US); e.g. application of an electric shock) is repeatedly preceded by a neutral event (conditioned stimulus (CS); e.g. presentation of a geometric shape; Lonsdorf et al., 2017). Over time, CS and US become associated - the CS is identified as a threat signal and evokes a fear response in anticipation of the US (conditioned response; CR), often measured as a rapid change in physiological functioning (e.g. increase in SCR or pupil dilation) following onset of the CS. Often, not one but two CS are included in fear conditioning tasks: one that is repeatedly paired with the US (denoted as the CS+) and another that is never paired with the US (CS-). Conditioned responses to the CS- are marked by the inhibition of fear and can be understood as safety responses (Laing & Harrison, 2021). Differential fear conditioning compares responses to CS+ and CS- and allows to draw inferences on both fear and safety learning while considering inter-individual differences in overall physiological responses (Lonsdorf et al., 2017).

The underlying neural substrates of fear conditioning have repeatedly been investigated in fMRI studies demonstrating a fear network that is activated following CS+ presentation (Fullana et al., 2016).

This network consists of several brain regions including the anterior insula, the ventral striatum, and the dACC. Following CS- presentation, a safety network is activated that includes cortical regions such as the primary somatosensory cortex (S1), vmPFC, and the posterior cingulate cortex (PCC; Fullana et al., 2016). Critically, fear and safety networks modulate each other in antagonistic fashion: the activation of the fear network inhibits the activation of the safety network and vice-versa.

An extensive animal literature shows that successful fear conditioning is dependent on opioidergic circuits (McNally et al., 2011). Mechanistically, this can be explained through the involvement of μ-opioid receptor signalling in prediction error (PE) learning – that is, the continuous updating the value of a stimulus by comparing predicted outcomes to actual outcomes (S. S. Y. Li & McNally, 2014). Differential fear conditioning can be thought of as PE learning, where, initially, both CS start out with a neutral value, since it is unknown which is associated with the US. Pairing one CS with the US creates a negative PE (i.e. outcome is worse than expected), which is used to update the CS value. Is this CS encountered again, US expectancy will be higher than before, as its aversive value has been increased. When a CS is followed by the absence of the US, this leads to a positive PE (i.e. outcome is better than expected), reducing the aversive value of the CS. The relative aversive value of the CS then determines the CR as either fear-inducing or fear-inhibiting. Prior research has demonstrated that learning from negative PEs is inhibited by high opioidergic tone, which is why acquisition of fear is attenuated after administering μ-agonists (Good & Westbrook, 1995) and potentiated after administering antagonists, both in animals and in humans (Eippert et al., 2008; Fanselow & Bolles, 1979). Learning from positive PEs on the other hand, is facilitated by high opioidergic tone (McNally & Westbrook, 2003), which is critical for fear extinction where the CR disappears when the CS+ is repeatedly not being followed by the US. Endogenous opioid receptor function can be downregulated following chronic opioid use, reducing opioidergic tone (Zubieta, 2000). This should lead to reduced learning from positive and increased learning from negative PEs, resulting in increased fear and decreased safety learning. Evidence of potentiated fear acquisition (Pennington et al., 2020) along with impaired fear extinction (Gu et al., 2008) in rodents subjected to chronic opioids supports this idea.

The only study to date that has investigated fear conditioning following chronic opioid use in humans found that opioid-addicted patients in OMT did not show increased but rather impaired fear conditioning, indicated by no SCR discrimination between CS+ and CS- (Basden et al., 2016). When prompted for contingency awareness (i.e. knowledge of which CS predicted the US), only 33% of participants in the opioid-addicted group correctly identified the correct stimulus, compared to 87% of healthy participants, which the authors propose could be explained by general cognitive deficits in patients (Basden et al., 2016). As these first results disagree with findings from the animal literature, many open questions remain regarding fear conditioning following chronic opioid use, such as whether

fear conditioning is differentially altered in addicted and non-addicted patients and whether changes in fear conditioning constitute a consequence or rather a pre-disposing factor for chronic opioid use.

Fear reversal

In an ever-changing world, it is not sufficient to once acquire knowledge about threat and safety, but rather it is necessary to constantly monitor and flexibly update this knowledge. When a previously safe stimulus (CS-) is surprisingly followed by a US, its value needs to be promptly updated (indicated by the expression of a CR) in order to adaptively change behaviour. Conversely, when a CS+ ceases to be followed by the US, the CR is gradually extinguished: the threat signal turns into a safety signal. Updating of fear and safety contingencies can be experimentally assessed using fear reversal tasks (Schiller et al., 2008). These paradigms consist of two phases, the acquisition phase, equivalent to a differential fear conditioning paradigm, and the reversal phase, where the CS-US associations are interchanged: without instruction, fear and safety contingencies switch so that the CS that previously preceded the US is now safe (old CS+/new CS-), while the formerly safe stimulus is now followed by the US (old CS-/new CS+). This reversal leads to an adjustment of the CR over the ensuing trials as the new CS+ elicits the CR, while the response to the new CS- extinguishes.

In the brain, the vmPFC has been repeatedly demonstrated to be particularly critical for safety updating, possibly by inhibiting previously learned threat associations (Schiller et al., 2008). As mentioned above, opioidergic transmission affects fear learning by mediating positive and negative PEs and is therefore likely involved in fear reversal. However, fear reversal was also found to be impaired following depletion of serotonin (Kanen et al., 2021), a neuromodulator involved among others in aversion processing and cognitive flexibility (Crockett & Cools, 2015; Matias et al., 2017), which indicates that non-opioidergic receptor function can also modulate fear and safety learning.

Fear conditioning and reversal in opioid addiction

Fear conditioning and reversal have been proposed as an important mechanisms in the maintenance of opioid addiction. Opioid addiction is defined by a chronically relapsing pattern of opioid use (Strang et al., 2020) thought to be motivated by avoiding withdrawal, which is caused by an allostatic state of the opioid system upon cessation of use (Koob, 2020). However, patients addicted to opioids often relapse after prolonged abstinence or while undergoing OMT (Hser et al., 2015), where homeostasis of the opioid system is regained or at least momentarily secured, nullifying the risk of withdrawal and - in theory - extinguishing motivation for use. Crucially, exposure to certain environmental cues can lead to the onset of opioid withdrawal. In a seminal study, experimenters repeatedly induced

withdrawal in opioid-addicted patients in OMT by administering naloxone and paired administration with a distinct high-frequency tone and odour (O'Brien et al., 1977). Following pairing of the tone/odour and naloxone-induced withdrawal symptoms, presentation of the tone alone sufficed to evoke both emotional and somatic symptoms of opioid withdrawal (conditioned withdrawal). Fear conditioning can explain this phenomenon and help to shed light on why long-term abstinent patients often relapse when encountering drug-related environments (Pantazis et al., 2021): Withdrawal symptoms serve as highly aversive US; previously innocuous stimuli (e.g. a friend's apartment where a withdrawal episode took place) take on threatening properties through their association with the US and become a threat signal (CS+). Presentation of the CS+ then leads to a conditioned fear response which in turn motivates opioid use, leading to relapse and restarting the addiction cycle. Critically, following prolonged abstinence or when undergoing OMT, fear and safety contingencies should be updated: No withdrawal is to be expected when re-encountering the environment, therefore the conditioned fear response should be extinguished. Impairments to safety updating in patients with opioid addiction could therefore help to explain why exposure to withdrawal-associated cues continues to elicit fear, and hence motivates relapse.

To date, no study has investigated fear reversal learning in patients with opioid addiction. In face of the key role of the vmPFC for safety updating (Schiller et al., 2008), it bears mentioning that this brain region has reduced grey matter volume in patients with opioid addiction (Wollman et al., 2017), which furthermore appears to be specific to opioid addiction when compared to cocaine addiction (Ceceli, Huang, et al., 2023). Interestingly, in a different study, vmPFC volume of currently abstinent opioid-addicted patients did not differ from healthy control participants (Tolomeo et al., 2016), which could either suggest that a smaller vmPFC is caused by chronic opioid use and recovers following abstinence or that patients with higher vmPFC volume have more success in ceasing opioid use. Deficiencies in vmPFC structure or function could underlie impaired safety updating and therefore contribute to conditioned withdrawal-related relapse. This remains, however, speculative as no study has tested the relationship between fear reversal, chronic opioid use, and opioid addiction.

Taken together, fear conditioning and reversal represent aspects of aversive processing essential to adaptive emotional responding. While fear conditioning figures to be altered following changes in opioidergic tone following chronic use, impaired safety updating has been suggested to contribute to the chronically relapsing nature of opioid addiction.

Aims & Hypotheses

In the following, I aim to characterise fear conditioning and reversal as well as their neural substrates following chronic opioid use in patients with and without opioid addiction. I hypothesise that fear conditioning is potentiated in the patients with chronic opioid use, indicated by a larger stimulus discrimination between the CS+ and the CS- compared to healthy participants. Concomitantly, I hypothesise that patients show hyper-activation of brain regions that are part of the fear network (anterior insula, ventral striatum, dACC) for the CS+ > CS- contrast, compared to healthy participants. For the fear reversal, I hypothesise that safety updating is impaired in the opioid-addicted patients, indicated by reduced stimulus discrimination between the new CS+ and the new CS- following reversal compared to the non-addicted patients. I further hypothesise that this effect is accompanied by reduced neural safety updating for the opioid-addicted patients, indicated by a lack of vmPFC activation for the new CS- > new CS+ contrast following fear reversal, while the vmPFC is activated in the non-addicted patients.

5.2 - Methods

Sample

A subset of the 86 participants included in the overall study sample completed the fear reversal task described in this chapter. 21 participants were not suitable for MRI scanning (1 control, 13 opioid-addicted, 7 non-addicted). 5 participants (2 control, 3 non-addicted) could not be reached or declined a second study visit. Additionally, 6 participants aborted the MRI scan due to either scanner malfunction, back pain, claustrophobia, or severe coughing (1 control, 3 opioid-addicted, 2 non-addicted). One healthy participant reported falling asleep during the task and was subsequently excluded. In total, 53 participants were included in the analyses reported in this chapter: 33 healthy participants, 7 opioid-addicted patients, and 13 non-addicted patients.

Task description

A threat reversal paradigm was applied to measure participants' physiological and neural responses to threatening and safe stimuli with reversing contingencies (Apergis-Schoute et al., 2017) inside an MRI scanner. In this task, participants were instructed to carefully observe pictures of angry male faces presented to them. Participants were further informed that they might receive electrical shocks to one of their wrists, but it was not specified when this would be the case. Two pictures of different men displaying an angry facial expression - overlaid with either green or purple colour to increase

discriminability - were shown in a pseudorandomised order on the screen for 4 s, with 12 s inter-trial-interval displaying a white-on-black fixation cross (see Figure 4). During the acquisition phase (trials 1 to 40), one of the two faces (Face A, the initial CS+) was partially reinforced with an electric shock (US), while the other face was never paired with a shock (Face B, initial CS-). The CS+ was presented 8 times with a shock delivered at cue onset and 16 times without shock while the CS- was presented 16 times. Throughout the task, we ensured that a CS+ presentation with shock would never be followed by another shock trial and that there would never be more than two consecutive trials of the same kind. Following the acquisition phase, the CS-US contingencies were reversed (reversal phase, trials 41 to 80), so that the face previously paired with a shock was now safe and the previously safe face was now intermittently reinforced with shocks. Again, the new CS+ (Face B, old CS-) was presented 8 times with a shock and 16 times without shock while the new CS- (Face A, old CS+) was presented 16 times. The reversal occurred uninstructed and without break, so participants were not aware that task parameters had changed. Images assigned to faces A and B were randomised between participants.

As this task was carried out following the habit learning task, where one Wasp electrode remained connected to one wrist while the other one was removed (see Chapter 8). Participants completed this task with the one electrode connected to either their right or left wrist (randomly assigned). Individual shock intensity was determined beforehand using a standardised shock work-up procedure (see Chapter 3). Before the fear reversal task started, we applied one shock to participants' wrist and participants confirmed that the shock was still aversive. During this task, no behavioural responses were carried out rendering skin conductance responses to the stimuli the primary outcome measure (see below).

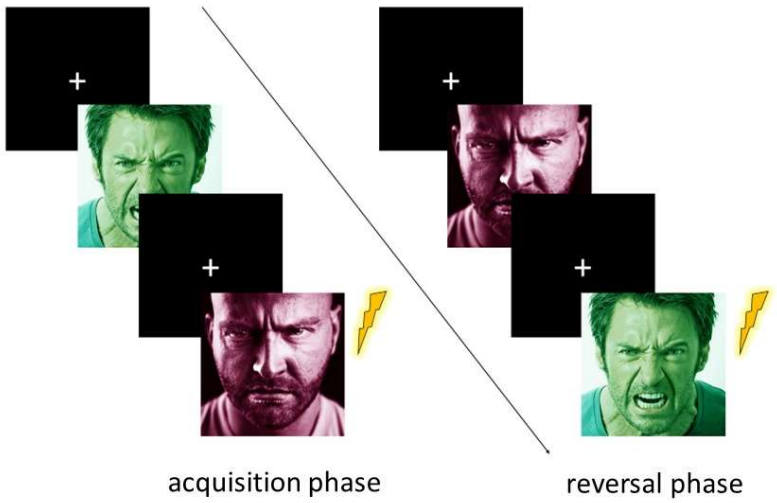


Figure 4. Schematic representation of the fear reversal paradigm.

Physiological recording, pre-processing and analysis

Similar to the original paper to use this task (Apergis-Schoute et al., 2017), acquisition and reversal phase were divided into an early and late stage each. During the course of the task, EDA was recorded from participants' left palm. EDA recording, visual inspection, pre-processing, and SCR computation were carried out as described in Chapter 3. For the analyses reported in this chapter, SCR data were averaged by stage (early acquisition (trials 1-20); late acquisition (trials 21-40); early reversal (trials 41-60); late reversal (trials 61-80) and stimulus (Face A; Face B, shock). All SCR data reported here were standardised and normalised (i.e. z-scored and log-transformed). Stimulus discrimination was computed for each stage by subtracting the mean SCR to Face B from the mean SCR to Face A, whereby only trials without shock were included, and used as primary outcome variable. Group differences in stimulus discrimination were tested using a 4x3 repeated measures ANOVA, with stimulus discrimination as dependent variable, stage (early acquisition/late acquisition/early reversal/late reversal) as within-subject factors, and group (control/non-addicted/addicted) as between-subject factor. For hypotheses testing, I tested the group-by-stage interaction using planned contrasts. Contrast weights were specified as follows: In Contrast 1, the opioid-using groups were combined and compared to the healthy participants for the early acquisition stage (group contrast weights: 1 -.5 -.5); in Contrast 2, the non-addicted patient group was compared to the addicted opioid-using group for the early acquisition stage (group contrast weights: 0 1 -1). For the early and late reversal stages, the same group comparisons were carried out, but for the inverse discrimination (Face B – Face A; group contrast weights: -1 .5 .5 and 0 -1 1). SCR responses to the US were analysed in a separate analysis, again using 4x3 repeated measures ANOVA, with US response as dependent variable, stage (early acquisition/late acquisition/early reversal/late reversal) as within-subject factor, and group (control/non-addicted/addicted) as between-subject factor.

Neuroimaging acquisition, pre-processing, and analysis

Acquisition and pre-processing of fMRI data were carried out as described in Chapter 3. For the first level (i.e. subject-level) analysis, individual general linear models were set up for each participant with one regressor for every stimulus during every phase of the task (12 regressors in total; CS+/CS-/US trials in early/late acquisition and early/late reversal). 4-second boxcars at stimulus onset were convolved with a canonical hemodynamic response function. Additionally, realignment parameters as well as their first and second derivatives were included as nuisance regressors to control for artefacts caused by head motion. To restrict analysis to grey matter BOLD responses, two additional nuisance regressors were added including white matter and CSF maps estimated during the segmentation in pre-processing. For each participant individually, regressor weights (beta images) were estimated

using the pre-processed 4D-images. Beta images of the 12 regressors of interest were then warped into template space and smoothed using a 6 mm full-width half-maximum (FWHM) Gaussian kernel. For the second level (i.e. group-level) analysis, a flexible factorial design was set up, including one regressor for each of the participants (subject factor), one for each of the groups (group factor), and one for each condition (e.g. "Early_Acquisition_CSplus"), as well as a total of 36 regressors for the group x condition interaction. Statistical parametric maps for each of the 96 regressors were estimated and analysed using a priori specified regions-of-interest (ROI) analyses, which were conducted by applying small-volume correction masks for insula, striatum, dACC/preSMA, and vmPFC/mOFC included in the WFU PickAtlas Toolbox for SPM (version 3.0.5). T-contrasts for flexible factorial designs were specified to test for interaction effects between groups and conditions (Gläscher & Gitelman, 2008). Comparable to the SCR analysis, a first group of contrast was specified to test for differences between healthy participants and patient groups (group contrast weights: 1 -.5 -.5), a second group of contrasts testes for differences between the two patient groups (group contrast weights: 0 1 -1). Results were corrected based on family-wise error rates with α-levels set to .05.

5.3 - Results

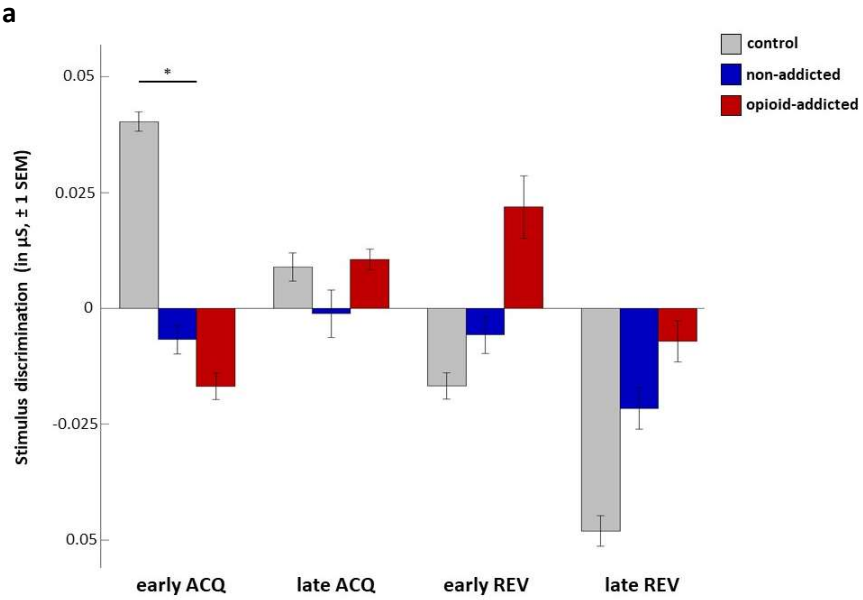
SCR data from 13 participants was excluded due to noisy recordings or non-responses to the US (8 control, 1 opioid-addicted, 4 non-addicted). The first 7 participants (all from the control group) included in the study were scanned using sub-optimal fMRI parameters, leaving to the exclusion of these subjects for neuroimaging analysis; however, their SCR data was included. One participant from the control group was excluded from the fMRI dataset due to excessive head motion (> 10 mm); SCR data from this participant was retained. The final fMRI sample consisted of 45 participants (25 control, 7 opioid-addicted, 13 non-addicted); the final SCR sample consisted of 40 participants (25 control, 6 opioid-addicted, 9 non-addicted). Groups did not differ in regard to the wrist the shock electrode was connected to ($\chi^2_2 = 2.2$, $p = .335$). Individual shock intensities applied for the task did not differ between the groups ($F_{2,53} = 0.7$, $p = 0.508$).

Physiological results

Repeated measures ANOVA revealed no significant main effect of group ($F_{2,37} = 0.4$, $p = .702$), main effect of stage ($F_{3,111} = 1.5$, $p_{corr.} = .218$) or group-by-stage interaction ($F_{6,111} = 1.7$, $p_{corr.} = .154$) on stimulus discrimination. In the early acquisition stage, planned contrast showed a significant difference between the patient and control groups ($t_{37} = 3.5$, $p = .001$, $d = 0.86$), indicating higher stimulus

discrimination in the control group (see Figure 5). Follow-up one-sample t-tests demonstrated that stimulus discrimination significantly differed from zero in the control group (t_{20} = 3.5, p = .002, d = 0.76), but not in the opioid-addicted (t_5 = -2.3, p = .066) or non-addicted (t_8 = -0.7, p = .503) group. Post-hoc tests revealed no group differences for the late acquisition stage (all $p_{corr.}$ > .800), where 95% confidence intervals (CIs) included zero for all three groups. Planned comparisons for the early reversal stage showed no significant difference between the two patient groups (t_{37} = 0.9, p = .397) or between the patient and control group (t_{37} = 1.2, p = .228). Stimulus discrimination for this stage differed from zero in the control group (t_{24} = -2.1, p = .045, d = 0.47), but not in the opioid-addicted (t_5 = 1.3, p = .244) or non-addicted group (t_8 = -0.5, p = .646). During the late reversal stage, no significant group differences between the two patient groups (t_{37} = 0.4, p = .694) or between the patients and the healthy participants were found (t_{37} = 1.5, p = .151). Again, stimulus discrimination did differ from zero for the healthy participants did t_{24} = -2.9, p = .007, d = 0.22), but not for the opioid-addicted (t_5 = -0.6, p = .545) or non-addicted patients (t_8 = -1.6, p = .148).

A separate repeated measures ANOVA analysing SCR to the US found a significant main effect of stage ($F_{3,111}$ = 14.3, $p_{corr.}$ < .001, η_p^2 =.28), but no main effect of group ($F_{2,37}$ = 2.0, p = .145) or group-by-stage interaction ($F_{6,111}$ = 1.6, $p_{corr.}$ = .184). Including shock intensity as a covariate into the analysis did not affect the results.



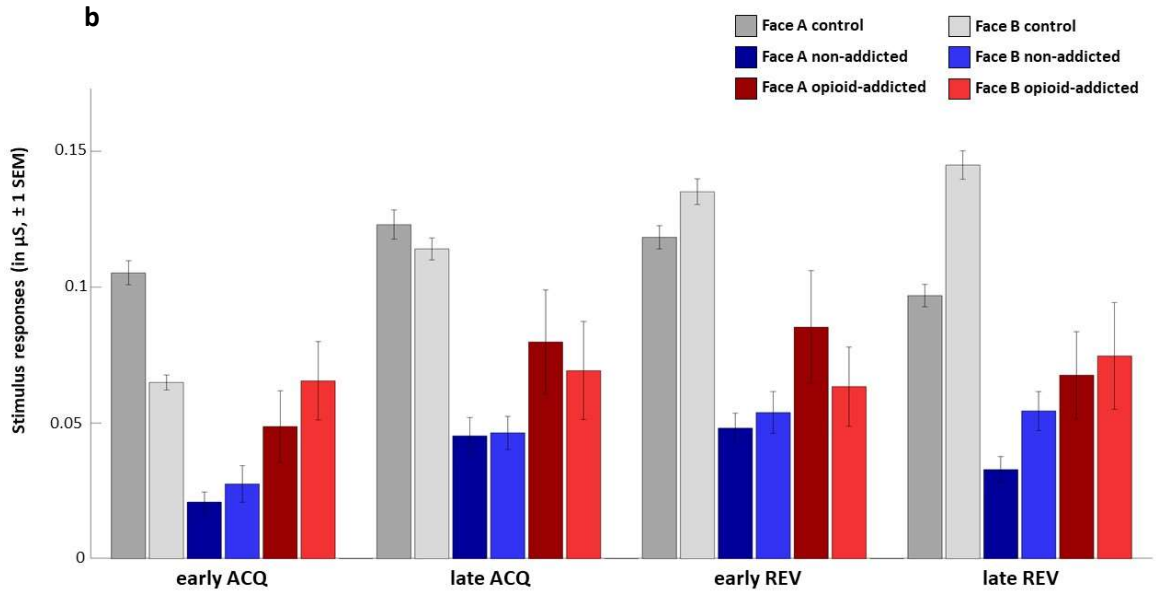


Figure 5. SCR stimulus discrimination (a) and stimulus responses (b) for all groups and stages. For better interpretability, figures display raw SCR data. For analyses, standardised and normalised data was used. ACQ: Acquisition phase; REV: Reversal phase.

Neuroimaging results

For the CS+ > CS- contrast in the early acquisition phase, ROI analysis revealed significantly higher activation for the patient groups compared to healthy participants in the left posterior insula (xyz = -39 -6 8, T = 5.1, p_{FWE} = .003). Conversely, healthy participants displayed significantly higher activation in the left caudate (xyz = -12 0 18, T = 5.6, p_{FWE} < .001) and left preSMA (xyz = -20 20 57, T = 6.7, p_{FWE} < .001). Figure 6 displays early acquisition CS+ and CS- activation in the significant clusters identified in the ROI analyses for all three groups. Exploratory whole-brain analysis for this contrast showed additional significant clusters at p_{FWE} < .05, including higher activation for controls in the mOFC and higher activation for patients in the right S1, left amygdala, and left vmPFC.

For the new CS- > new CS+ contrast during early reversal, ROI analysis revealed no significant clusters when comparing the two patient groups, at either p_{FWE} < .05 or $p_{uncorr.}$ < .001. During the late reversal stage, however, opioid-addicted patients showed higher activation compared to non-addicted patients in a vmPFC cluster (xyz = 2 32 -26, T = 4.8, p_{uncorr} < .001), albeit only after lowering the significance level $p_{uncorr.}$ < .001 (see Figure 7). No voxels survived FWE-correction for exploratory whole-brain analysis.

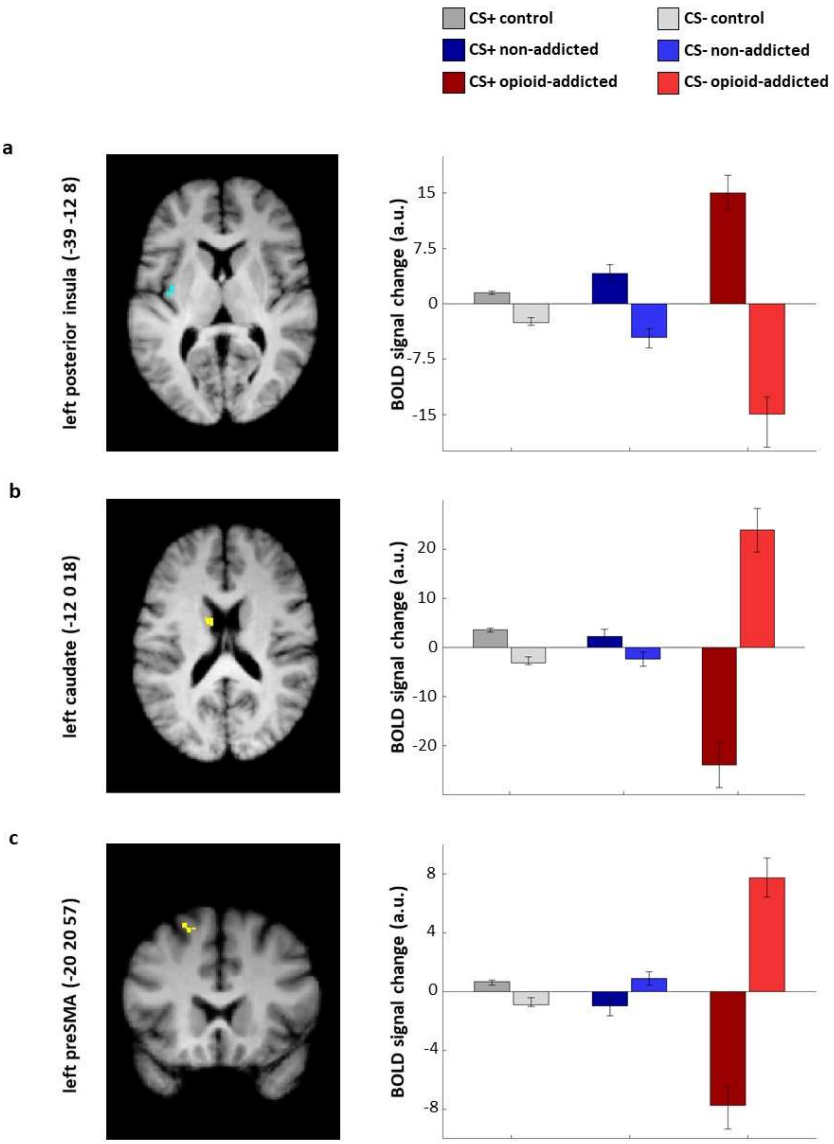


Figure 6. Significant activations in pre-defined ROIs for the CS+ > CS- contrast during early acquisition. Clusters are presented at $p_{FWE} < .05$. Yellow denotes higher activations for healthy participants, cyan denotes higher activations for the patient groups. BOLD activation is averaged across all voxels of the cluster.

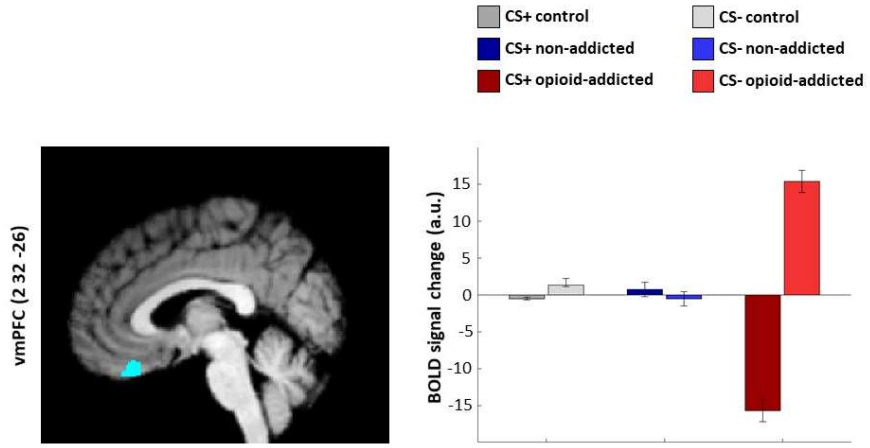


Figure 7. Neural safety signalling following reversal. Significant activations in the pre-defined ROI for the new CS- > new CS+ contrast during late reversal. Cluster is presented at $p_{uncorr.} < .001$. Cyan denotes higher activations for opioid-addicted compared to non-addicted patients. BOLD activation is averaged across all voxels of the cluster.

Whole-brain analysis for the main effect of US yielded significant clusters in the right anterior insula (xyz = 33 26 4, T = 6.9, p_{FWE} < .001), left anterior insula (xyz = -33 24 6, T = 6.2, p_{FWE} < .001), right preSMA (xyz = 2 12 51, T = 5.0, p_{FWE} = .020), and occipital cortex (xyz = 15 -84 -9, T = 10.1, p_{FWE} < .001). No group differences were found in US responses when comparing healthy participants to the patient groups or when comparing the two patient groups in whole-brain analyses thresholded at either p_{FWE} < .05 or $p_{uncorr.}$ < .001.

Relationship between physiological and neural responses

Associations between SCR discrimination during early acquisition, late reversal, and US presentation were correlated with neural activations extracted from significant clusters stemming from ROI analysis; Spearman correlation was used to adjust for extreme values in neural activation patterns. For healthy participants and non-addicted patients, SCR discrimination during early acquisition did not significantly correlate with stimulus discrimination in any ROI. For the opioid-addicted patients, early acquisition SCR discrimination correlated positively with early acquisition neural discrimination in the left caudate (ρ = .89, p = .033). During late reversal, SCR discrimination was not significantly correlated with vmPFC discrimination in either group. In the non-addicted patients only, SCR US responses during early acquisition positively correlated with early acquisition neural stimulus discrimination in the posterior insula (ρ = .77, p = .021).

Associations with opioid use characteristics

For opioid-addicted patients, duration of opioid use was significantly and negatively correlated with US responses in the left anterior insula (ρ = -.82; p = .034), while showing a trend in the right anterior insula (ρ = -.71; p = .088). Current daily opioid dose was not correlated with any physiological or neural responses. For the non-addicted patients, duration of opioid use was significantly correlated with SCR stimulus discrimination in early acquisition (ρ = -.67; p = .047), and early reversal (ρ = -.84; p = .005), while current daily opioid dose was significantly associated with early acquisition stimulus discrimination in the left preSMA (ρ = .62; p = .024) and with US responses in left (ρ = -.66; p = .015) and right anterior insula (ρ = -.60; p = .032).

5.4 - Discussion

In this chapter, my aim was to characterise fear conditioning and reversal following chronic opioid use while also analysing the neural basis of these mechanisms. Patients with and without opioid addiction as well as healthy participants completed a fear reversal paradigm, where physiological (SCR) and neural (fMRI BOLD) responses to threatening and safe stimuli were measured during fear acquisition and reversal.

Contrary to my first hypothesis, fear conditioning measured by SCR stimulus discrimination in the early acquisition stage was impaired rather than potentiated in patients with chronic opioid use. While healthy participants learned to discriminate between the threatening and the safe stimulus, patient groups failed to do so, indicated by no difference between the conditioned responses to the stimuli. Importantly, the lack of stimulus discrimination cannot be explained by alterations in the processing of the US itself as all groups showed similar SCR on trials in which the electric shock was presented. This suggests that aversive learning and not the initial encoding of aversive stimuli is impaired following chronic opioid use, in accordance with the findings of the only other study investigating fear conditioning in patients in OMT (Basden et al., 2016), which found no differential fear conditioning despite intact US responses.

Impaired differential fear conditioning in the patient groups was mirrored by inconsistent and intertwined activation of neural fear and safety networks. While activation patterns in some parts of the fear network, such as the anterior insula and dACC, were similar to healthy participants, patients showed activation in regions included in the safety network in response to the threatening stimulus, such as the primary somatosensory cortex and the vmPFC (Fullana et al., 2016). In response to the safe stimulus, patients hyper-activated multiple brain structures central to the fear network (Fullana et al., 2016), such as the caudate, putamen, and preSMA. Interestingly, patients also showed increased activation to the threatening compared to the safe stimulus in the posterior insula, a brain region which encodes the intensity of incoming painful stimuli (Segerdahl et al., 2015). This suggests that on some level stimulus discrimination did indeed occur in the patients; however, disorganised overall activation of neural fear and safety networks appears to impede aversive learning.

These neuroimaging findings allow to draw inferences on the mechanisms underlying impaired fear conditioning in patients. Successful fear learning hinges on assigning different aversive values to the threatening and safe stimulus. Initially, values of the two are identical; as soon as one stimulus is followed by an aversive event, negative prediction error signals are utilised to increase the aversive value of that stimulus. Inconsistent activation of the fear network, particularly of the caudate where aversive prediction error signals are encoded (Seymour et al., 2007), prevents the continuous updating of aversive values required for fear learning. The lack of physiological discrimination between

conditioned responses to the stimuli observed in this chapter could therefore be the result of deficient striatal prediction error learning. The correlation between caudate activation and SCR stimulus discrimination in the opioid-addicted patients underscores the notion that impaired processing of aversive prediction errors leads to impaired fear learning in addiction. Notably, this relationship was absent in the non-addicted patients, which could imply that impaired fear conditioning in this group might be caused by something other than impaired striatal prediction error signalling. In this group, longer duration of opioid use was correlated with smaller SCR stimulus discrimination, possibly indicating that impaired fear learning is caused by chronic opioid use in the non-addicted group, but independent from chronic use in addicted patients. However, larger patient samples would be required to reliably dissociate mechanistic impairments underlying a lack of differential fear conditioning in these two groups.

My second hypothesis predicted impaired safety updating in patients addicted to opioids compared to non-addicted patients. Since no fear conditioning occurred in these two groups, it was not possible to investigate the updating of fear and safety contingencies; after all, one can only re-learn what has been learned to begin with. While healthy participants showed the expected pattern (larger SCR for the new CS+ compared to the new CS- following reversal) similar to other studies using the same paradigm (Apergis-Schoute et al., 2017; Roberts et al., 2022), stimulus discrimination did not differ from zero in either patient group following the reversal procedure. Unexpectedly, opioid-addicted patients showed increased activation to the new CS- in the vmPFC compared to the non-addicted patients during late reversal, suggesting that safety signalling occurred in this group but not in the non-addicted patients. However, activation in this cluster was not associated with SCR discrimination which suggests that other neural processes might have impeded expression of adequate conditioned fear and safety responses.

Lastly, it is notable that all three groups selected similar shock intensities during the pre-task shock work-up procedure and physiological and neural responses to electric shocks were comparable between groups. Previous research suggested reduced thresholds for aversive stimulation following chronic use (opioid-induced hyperalgesia; Higgins et al., 2019; Trøstheim & Eikemo, 2024). From this, lower shock intensities and higher US responsiveness could have been expected for the patient groups. At the same time, all patients were using opioids at the time of testing, which might have counteracted increased pain perception and normalised neural responses to aversive stimulation (Lee et al., 2014). In support of this idea, a higher current opioid dose was associated with reduced US responses in the anterior insula for the non-addicted patients. Such a relationship was absent in the opioid-addicted patients, and it is worth mentioning that methadone dose has previously been found to be positively rather than negatively correlated with SCR to electric shocks in opioid-addicted patients (Basden et al.,

2016). Instead, in opioid-addicted, but not non-addicted, patients the duration of opioid use was negatively associated with anterior insula US responses, which could indicate long-term opioidergic adaptations following chronic use. However, this correlation is based on data of only six individuals and therefore requires replication in larger samples. Hence, the relationship between pain processing on the one hand and chronic and acute opioid use on the other remains elusive.

In this chapter, fear conditioning was found to be impaired in patients with chronic opioid use, irrespective of whether they were addicted to opioids or not and despite intact processing of the aversive unconditioned stimulus. Neuroimaging results indicate overlapping hyper- and hypo-activation of neural fear and safety networks, pointing towards disorganised neural responses to threatening and safe stimuli underlying the learning deficit. Different relationships between fear conditioning outcomes and opioid-use characteristics between the two opioid-using patient groups suggest differential impairments resulting in fear learning deficits; nevertheless, the small sample sizes of the patient groups in this chapter somewhat limit the interpretability of the relationship between fear conditioning and chronic opioid use.

Chapter 6 - Aversive instrumental learning

6.1 - Introduction

In addition to the Pavlovian fear conditioning covered in the previous chapter, learning from aversive experiences is also mediated by aversive instrumental learning, which is related but yet separate to fear conditioning. By way of aversive instrumental learning (also called instrumental avoidance learning), we learn which behaviours to perform to avoid aversive events. As argued in Chapter 1, avoidance and opioid use are intricately linked: While avoidance of aversive states such as pain, stress or anxiety is the primary motivation for opioid use (Han et al., 2017), avoidance of withdrawal and negative affect is considered the driving motivational force behind the maintenance of opioid addiction (Koob, 2020). As the transition from initial opioid use to opioid addiction can be conceptualised as an escalating spiral of avoidance behaviour, it is crucial to investigate instrumental avoidance learning as potential mechanism in this development.

Instrumental avoidance learning can be understood within the framework of reinforcement learning. Based on Thorndike's (1927) *law of effect*, instrumental learning assumes that the frequency or probability of performing a behaviour is determined by the outcome it has led to in the past (Sutton & Barto, 2018). When a behaviour led to a desirable outcome, it is reinforced and the likelihood of repeating the behaviour increases. Conversely, when a behaviour led to an undesirable outcome, probability of repetition decreases. A desired outcome could either consist of obtaining a reward (positive reinforcement) or avoiding punishment (negative reinforcement). Negative reinforcement is the mechanism suggested to underlie instrumental avoidance learning: behaviours that result in aversive outcomes acquire high aversive values and are therefore reduced or ceased entirely, while competing behaviours resulting in the avoidance of the outcome acquire low aversive values and are thus selected more frequently. In experimental settings, aversive instrumental learning tasks commonly present participants with multiple initially neutral stimuli, asking them to select one of them (Krypotos, 2015). If participants select an incorrect option, punishment occurs (e.g. electrical shock, money loss), whereas a correct selection results in the avoidance of punishment. Such action-outcome contingencies can be either deterministic (i.e. selection of the correct option always leads to avoidance) or probabilistic (i.e. selection of the correct option increases the probability of avoidance). Over the course of the experiment, participants accumulate evidence to select behavioural options associated with a lower probability of punishment, indicating successful avoidance learning. In the brain, feedback loops between multiple parts of the frontal cortex and the striatum provide the neural basis for successful avoidance learning. Punishment and avoidance PEs are encoded in different sub-regions of the striatum, such as caudate, NAcc, and posterior putamen (Delgado et al., 2008; Eldar et al., 2016; H. Kim et al., 2006). The expected outcome values of the available behavioural options

(encoded in the anterior insula) are continually updated by integrating punishment and avoidance PE signals from the striatum. In successful avoidance learning, the behaviour associated with the lowest aversive outcome value is selected from several competing behavioural options, which activates the mOFC/vmPFC, implicated in comparing values of several options to each other, and the striatum, crucial for action selection (Palminteri et al., 2012; Pessiglione et al., 2006).

Chronic opioid use can be conceptualised as learned avoidance behaviour. Initial opioid use leads to the avoidance of negative states (e.g. relief from physical pain, stress), which serves as a negative reinforcement strengthening the action-outcome association. Given this critical role of avoidance for chronic opioid use, alterations to avoidance learning could influence trajectories of opioid use. To date, only very few studies have investigated aversive instrumental learning following chronic opioid use and, so far, only in opioid-addicted patients. In one study, patients showed comparable avoidance behaviour to healthy participants (Gradin et al., 2014) in a probabilistic learning paradigm; however, neither group displayed avoidance accuracy above chance, indicating no avoidance learning regardless of group. Unsuitable task parameters (e.g. low number of trials, small difference in avoidance probabilities assigned to the correct and incorrect options) are likely to explain the lack of significant avoidance learning (see Pessiglione et al. (2006), where the expected effect is found in the same task with different parameters). While this prevented detecting potential learning impairments in behaviour, patients differed from controls in their neural responses: In contrast to healthy participants, patients failed to show activation relating to the encoding of avoidance PEs in the ventral striatum and demonstrated reduced insula activation, indicating weaker representation of expected outcome values (Gradin et al., 2014). These results indicate that opioid-addicted patients do not show the pattern of neural activations associated with successful avoidance learning; however, given the lack of a functional experimental design, these results have to be interpreted with caution. Notably, reward learning was successful for both groups, indicating an avoidance-specific rather than general learning deficit. Another study revealed that, compared to healthy participants, opioid-addicted patients showed lower accuracy when learning from punishment but not when learning from reward (Myers et al., 2017). Here, the authors proposed that the aversive learning deficit could be caused by an undervaluation of aversive outcomes; however, the neuroimaging findings reported in Gradin et al. (2014) demonstrate neural responses to money loss are intact. Instead, patients display difficulties in value updating, which ties in with findings from Chapter 5, where fear conditioning was impaired in patients despite intact responses to the aversive stimulus. The influential two-factor learning theory (Mowrer, 1951) postulates fear conditioning to be an integral part of avoidance learning: punishment following a behaviour can be understood as a Pavlovian CS-US association, leading to fearful conditioned responses in anticipation of this behaviour. Fear relief by selecting a different behavioural option would then serve as negative reinforcement, enabling instrumental learning. Therefore, it is

possible that an inconsistent updating of aversive values, related to the erratic activation of striatal structures observed during fear conditioning, could underlie impaired instrumental avoidance learning in patients with chronic opioid use.

Difficulties in avoidance learning following chronic opioid use, however, do not explain while some individuals develop opioid addiction while others do not. However, it has been proposed that conditioned negative reinforcement (Pantazis et al., 2021) fundamentally contributes to the compulsive opioid use that characterises addiction. Referring to the notion of conditioned withdrawal presented in Chapter 5, this suggests that stimuli previously paired with highly aversive withdrawal symptoms induce a strong fear of withdrawal. Since the onset of withdrawal can be reliably avoided by opioid use, in this conceptualisation, opioid use can be understood as conditioned avoidance (Pantazis et al., 2021). Performance of avoidance behaviour can, in turn, serve to maintain the fearful conditioned response (see Krypotos (2015) for the bidirectional relationship between the two), which is why Pavlovian associations are not extinguished over time. This could explain while it is not uncommon that addicted patients relapse after prolonged abstinence. If this was the case, opioid-addicted, but not non-addicted patients, might experience a bias towards avoiding stimuli that they had learned to associate with aversive states such as withdrawal symptoms, subsequently showing increased avoidance behaviour when avoiding aversively conditioned stimuli.

Taken together, while preliminary evidence suggests impaired aversive instrumental learning following chronic opioid use, it remains uncertain if and to what extent such a deficit exists and whether or not it is specific to patients with opioid addiction. Moreover, so far it is unknown whether conditioned avoidance, the avoidance of stimuli previously paired with aversive stimuli, differs between the two patient groups.

Aims & Hypotheses

In this chapter, my aim is to characterise instrumental aversive learning following chronic opioid use. I will apply a novel two-part aversive learning paradigm designed to test instrumental avoidance learning while concomitantly investigating potential effects of conditioned avoidance. For this, instrumental learning will be preceded by a fear conditioning task in which some stimuli will be paired with aversive and some with neutral stimuli. These stimuli subsequently have to be avoided in the instrumental learning task, testing the influence of acquired fear responses on avoidance behaviour.

I hypothesise that instrumental avoidance learning is impaired following chronic opioid use, indicated by lower avoidance accuracy compared to healthy participants, while avoidance accuracy does not differ between patients with opioid-addiction and non-addicted patients. Furthermore, based on the

notion of increased conditioned avoidance in opioid addiction, I hypothesise that opioid-addicted patients, compared to non-addicted patients, show increased higher avoidance behaviour towards stimulus previously paired with aversive, compared to neutral, content.

6.2 - Methods

Sample

All 86 participants included in this study completed the two-part aversive learning task and were subsequently included for analysis in this chapter.

Behavioural task

The aversive learning task consisted of two parts. In the first part of the task (Pavlovian fear conditioning), coloured tartan patterns were followed by aversive or neutral stimuli while conditioned responses to the tartans were measured by SCR. In the second part (aversive instrumental learning), participants learned to avoid money loss in a probabilistic learning environment that included the previously conditioned tartans.

Prior to the start of the task, two electrodes were applied to the palm of the non-dominant hand to record EDA (see Chapter 3). For the first part of the task, participants were asked to carefully observe tartans displayed one after another on the screen and were instructed that tartans might be followed by a combination of an image and a sound, some of which could be perceived as disturbing or off-putting. To stay engaged during the task, participants were asked to count how many tartans were surrounded by a blue frame. Following the instructions, the task started. In every trial, a white-on-black fixation cross was followed by a coloured tartan (CS) and afterwards the presentation of a combination of a picture and an audio file (US; see Figure 8a). Nature of the US depended on the trial type (aversive or neutral): In aversive trials, two photos of human vomit were combined with audio files of a person throwing up ($US_{aversive}$). In the neutral trials, two photos of innocuous everyday objects (a shopping cart, coat hangers) were combined with generic ringing sounds ($US_{neutral}$). Two coloured tartans were randomly assigned to each trial type ($CS_{aversive}$ and $CS_{neutral}$), resulting in four CS-US pairings, which were presented to participants in random order. Either at the beginning or at the end of the task (counterbalanced), two additional tartans were presented in a block design. These trials served the purpose of familiarising the participants with the tartans without any associative learning involved, important for identifying the impact of conditioned avoidance in the second part of the task. Following the presentation of the familiar tartans ($CS_{familiar}$), no US was presented: the screen turned

black and no audio was played. Each tartan was presented 10 times, resulting in a total of 60 trials[6]. After the last trial participants were asked how many blue frames they had counted. Participants then had to rate the different visual stimuli as US for pleasantness and arousal on visual analogue scales from 0 to 100. For this first part of the task, primary outcome measures were SCR discrimination between the different trial types ($CS_{aversive}$ and $CS_{neutral}$) as well as the rating scales.

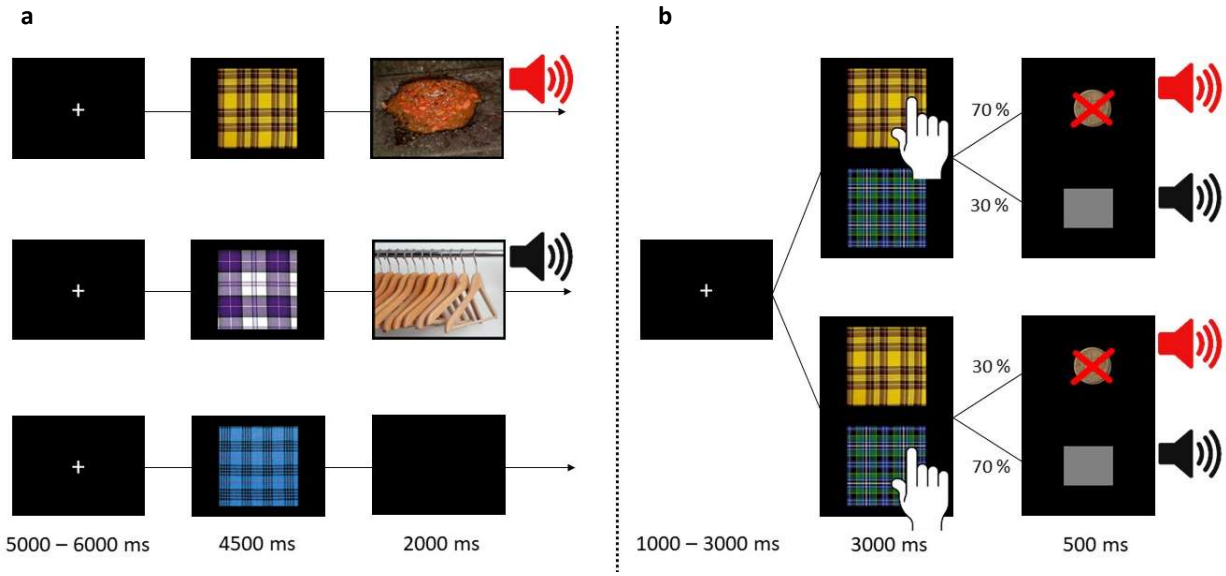


Figure 8. Schematic representation of the two-part aversive learning paradigm consisting of Pavlovian conditioning (a) and instrumental learning utilising the conditioned tartans from the first part of the task (b).

The second part of the task immediately succeeded the first part. Participants were instructed to retain as much as possible from a €50 virtual bank account by making correct decisions. They were informed that selection of some tartans was linked to a high probability of losing 50ct while selecting others was linked to a low probability. If participants failed to respond or responded late, they would lose €1. Every trial consisted of a white-on-black fixation cross, followed by two differently coloured tartans shown next to each other (see Figure 8b). Participants could select one tartan by touching it on the

[6] Task parameters pertaining to the number of trials as well as duration of CS presentation were not the ones initially used. After starting data collection, it was noticed that a relatively short presentation of the CS (1500 ms) prevented the analysis of CR to the CS as physiological response curves would overlap with responses to the US itself. Accordingly, the timing was adapted to the reported parameters while also reducing the number of trials from 120 to 60 to accommodate the elongated trial structure. 56 participants (28 healthy participants, 18 non-addicted patients, 10 opioid-addicted patients) completed the task before the first part was adapted, while 30 participants (10 healthy participants, 7 non-addicted patients, 13 opioid-addicted patients) completed the task with the final parameters, only of whom SCR data is included. Critically, neither of the behavioural outcome variables (US arousal and unpleasantness ratings, difference between number of actual and reported blue frames, number of trials without response, avoidance accuracy) differed between participants who completed either the final or the earlier version of the task, so that all participants were included in the analysis of these measures.

screen; afterwards, a combination of visual (a crossed-out 50ct coin to signal money loss or a neutral grey square to signal loss avoidance) and auditory feedback was provided (an aversive buzzing noise indicating money loss or a neutral ringing sound indicating loss avoidance). If participants did not respond within the 3 second response window, a crossed-out €1 coin was presented together with a different, highly aversive noise. Participants' current balance was at all times displayed underneath the two tartans. In every trial, one of the two displayed tartans had already been introduced during the first part of the task (either in aversive, neutral or familiarisation trials). The other tartan was one of six novel tartans. Selecting a tartan from the first part of the study was associated with a 70% chance of money loss, while selecting a novel tartan resulted in money loss in 30% of the trials, so that it was advantageous to avoid an old tartan in favour of a novel tartan. Each of the six familiar-novel combinations was presented 20 times (left-right alternated), resulting in a total 120 trials. After 60 trials, there was a short intermission and participants were free to continue after a short break. Primary outcome measure was avoidance accuracy, i.e. the percentage of novel tartans selected over old tartans, both overall and for all the tartan types (aversive, neutral or familiar) separately.

Psychophysiological recording, pre-processing, and analysis

Procedures pertaining to EDA recording and pre-processing are described in Chapter 3. SCR were computed as event-related signal change in a response window defined as 500 to 3000 ms following stimulus onset. Statistical analysis reported in this chapter are based on standardised and normalised (i.e. z-scored and log-transformed) SCR data. For each participant, SCR data for the two trial types (aversive and neutral) was averaged. Stimulus discrimination was computed by subtracting SCR to $CS_{neutral}$ from SCR to $CS_{aversive}$ and analysed using a one-way ANOVA, with stimulus discrimination as dependent variable and group (control/non-addicted/addicted) as independent variable. Planned contrasts were used for comparing healthy participants to patients (group contrast weights: -1 .5 .5) and opioid-addicted to non-addicted patients (group contrast weights: 0 1 -1). Group differences in SCR to the $US_{aversive}$ and $US_{neutral}$ were tested in a 2x3 repeated measures ANOVA with stimulus response as dependent variable, stimulus (aversive/neutral) as within-subject factor, and group (control/non-addicted/addicted) as between-subject factor.

Analysis of behavioural data

For the first part of the task, unpleasantness and arousal ratings for each of the four US were averaged within trial type and then compared separately between groups using a repeated measures ANOVA with rating score as dependent variable, stimulus (aversive/neutral) as within-subject factor, and group

(control/non-addicted/addicted) as between-subject factor. The reported number of tartans presented in a blue frame was subtracted from the actual number of blue frames; the difference was compared between groups in a one-way ANOVA with difference in reported blue frames as dependent variable and group (control/non-addicted/addicted) as independent variable. For the second part of the task, avoidance accuracy was computed as the proportion of trials on which the novel tartan was selected over any of the tartans from the first part of the study, both across stimulus types as well as for each stimulus (aversive, neutral, familiar) separately. To test learning over time, 40 trials per stimulus were divided into 4 bins of 10 trials each. Hypotheses were testing using a 3x4x3 repeated measures ANOVA with avoidance accuracy as dependent variable, stimulus (aversive/neutral/familiar) and bin (1/2/3/4) as within-subject factor, and group (control/non-addicted/addicted) as between-subject factor. For the second hypothesis, group differences in overall avoidance accuracy were tested using planned contrasts on the main effect of group, comparing healthy participants to both patient groups (group contrast weights: -1 .5 .5) and opioid-addicted to non-addicted patients (group contrast weights: 0 1 -1). For the third hypothesis, planned contrasts were defined for the group-by-stimulus interaction, testing avoidance accuracy on aversive trials over neutral trials for the group difference between the two patient groups (interaction contrast weights: 0 0 0 -1 1 0 1 -1 0). To further investigate avoidance learning, gains in avoidance accuracy between early and late learning ($bin_4 - bin_1$) were computed for each stimulus type and compared between groups (3x3 repeated measures ANOVA).

6.3 - Results

SCR datasets from 5 participants (2 healthy participants, 1 non-addicted patient, 2 opioid-addicted patients) had to be discarded because of noisy recordings or non-responsiveness to the US.

Part 1 – US pleasantness and arousal ratings

For the US pleasantness ratings, repeated measures ANOVA revealed a significant main effect of stimulus ($F_{1,83}$ = 144.0, p < .001, η_p^2 = .63), indicating higher pleasantness ratings for $US_{neutral}$ compared to $US_{aversive}$ (aversive: 25.0 ± 20.9; neutral: 59.5 ± 15.9). The group-by-stimulus interaction ($F_{2,83}$ = 3.3, p = .042, η_p^2 = .07) was also significant but no post-hoc test showed significant results following correction for multiple comparisons. The main effect of group was not significant ($F_{2,83}$ = 0.4, p = .668). For the US arousal ratings, repeated measures ANOVA showed a significant main effect of stimulus ($F_{1,83}$ = 21.6, p < .001, η_p^2 = .21), indicating higher self-reported arousal for $US_{aversive}$ compared to $US_{neutral}$, and a significant stimulus x group interaction ($F_{2,83}$ = 4.7, p = .012, η_p^2 = .10). Post-hoc tests revealed

that healthy participants and non-addicted patients rated the $US_{aversive}$ as more arousing than $US_{neutral}$ (control: t_{83} = 5.7, p_{corr} < .001, d = 1.2; non-addicted: t_{83} = 2.5, p_{corr} = .014, d = 0.64) but the opioid-addicted patients did not (t_{83} = 0.6, p_{corr} = .559). The main effect of group was not significant ($F_{2,83}$ = 1.1, p = .329).

Part 1 – Number of correct frames

The difference between the reported number of tartans in blue frames and the actual number of tartans in blue frames differed significantly between groups ($F_{2,83}$ = 6.2, p = .003, η_p^2 = .13), with post-hoc tests demonstrating that opioid-addicted patients showed a greater discrepancy between reported and actual frames compared to controls (t_{83} = 3.3, $p_{corr.}$ = .004, d = 0.88) and non-addicted patients (t_{83} = 2.9, $p_{corr.}$ = .014, d = 0.83).

Part 1 – SCR results

For the analysis of SCR discrimination between conditioned responses to $CS_{aversive}$ and $CS_{neutral}$, the main effect of group just failed to meet significance ($F_{2,24}$ = 3.0, p = .069). However, planned contrasts revealed significantly higher stimulus discrimination in the healthy participants compared to the patient groups (t_{23} = 2.4, p = .025, d = 1.0), while discrimination did not differ between the patient groups (t_{23} = -0.8, p = .441). One-sample t-tests demonstrated that stimulus discrimination was unequal to zero in the healthy participants (t_7 = 3.0, p = .020, d = 1.1), but not in either of the patient groups (opioid-addicted: t_{10} = -0.4, p = .714; non-addicted: t_6 = -1.4, p = .199), indicating impaired fear conditioning in patients. Repeated measures ANOVA to test for potential group differences in SCR to $US_{aversive}$ and $US_{neutral}$ revealed a significant main effect of stimulus ($F_{1,23}$ = 11.3, p = .003, η_p^2 = .33), but no main effect of group ($F_{2,23}$ = 2.2, p = .139) or stimulus-by-group interaction ($F_{2,23}$ = 0.5, p = .639). Post-hoc test revealed a higher SCR to the $US_{aversive}$ compared to the $US_{neutral}$ across groups (t_{23} = 3.4, p = .003, d = 0.88).

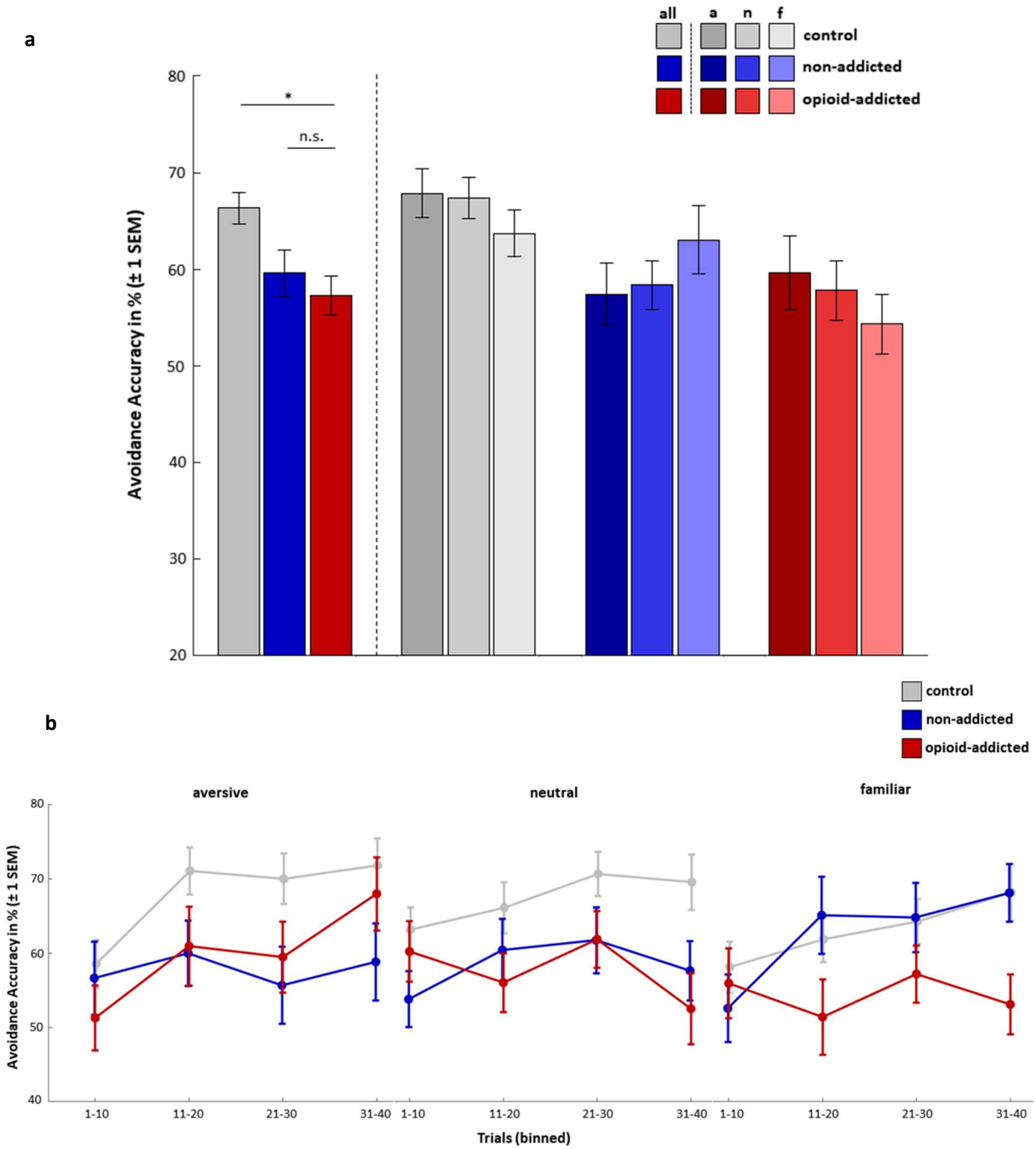


Figure 9. Avoidance accuracy by group and stimulus (a) as well as over trials (b). "a" refers to aversive, "n" to neutral, and "f" to familiar trials.

Part 2 – Avoidance accuracy

See Figure 9 for overall and stimulus-specific avoidance accuracy. Number of trials without response did not differ between groups ($F_{2,83}$ < .01, p = .993). Repeated measures ANOVA revealed significant main effects of group ($F_{2,83}$ = 5.7, p = .005, η_p^2 = .12) and bin ($F_{3,249}$ = 6.7, p < .001, η_p^2 = .07), while the main effect of stimulus ($F_{2,166}$ = 0.4, $p_{corr.}$ = .672) as well as interactions between group and bin ($F_{6,249}$ = 1.0, p = .440), group and stimulus ($F_{4,166}$ = 1.4, $p_{corr.}$ = .238), and bin and stimulus ($F_{6,498}$ = 1.7, p = .110) failed to reach significance. The three-way interaction between group, bin, and stimulus was also not significant ($F_{12,498}$ = 1.3, p = .206). Planned contrasts showed significantly higher overall avoidance accuracy for healthy participants compared to patients (t_{83} = 3.3, p = .001, d = 0.36), while

avoidance accuracy did not differ between the two patient groups (t_{83} = 0.7, p = .458). Planned contrasts for the group-by-stimulus interaction indicated a similar difference in avoidance accuracy on aversive trials compared to neutral trials between the two patient groups (t_{83} = 0.5, p = .643).

Exploratory analysis of differences in avoidance accuracy between bins 4 and 1 revealed a significant group-by-stimulus interaction ($F_{4,166}$ = 2.7, p = .031, η_p^2 = .06), with post-hoc tests demonstrating higher accuracy gains for aversive compared to neutral (t_{83} = 3.0, $p_{corr.}$ = .011, d = 0.93) and familiar (t_{83} = 2.7, $p_{corr.}$ = .016, d = 0.74) trials in the opioid-addicted group (see Figure 10). Additionally, opioid-addicted patients had lower accuracy gains on familiar trials compared to non-addicted patients (t_{83} = -2.5, $p_{corr.}$ = .047, d = -0.70). 95% confidence intervals included zero for the following group-stimulus pairings, indicating no significance difference between avoidance accuracy in bins 4 and 1:

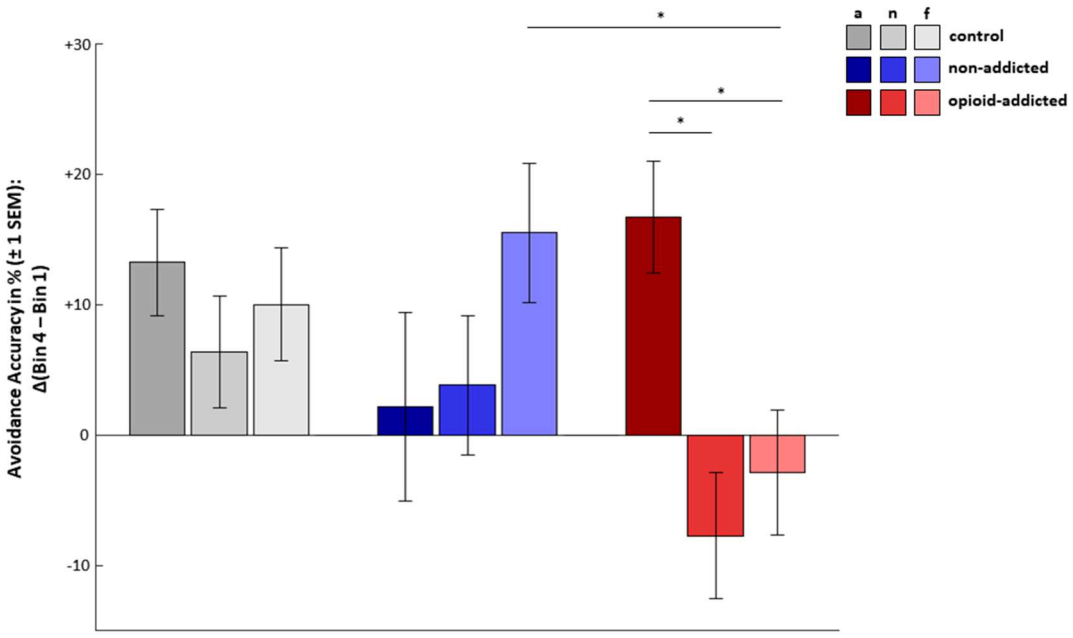


Figure 10. Gains in avoidance accuracy between early and late learning by group and stimulus. "a" refers to aversive, "n" to neutral, and "f" to familiar trials.

Opioid use characteristics and risk factors

For the opioid-addicted patients, neither SCR stimulus discrimination, SCR to either US, or avoidance accuracy on either trial type was significantly associated with current opioid dose, duration of opioid use, age of opioid use onset, or self-reported obsessive-compulsive drug use (OCDUS score). For the non-addicted patients, duration of opioid use was significantly and negatively correlated with SCR discrimination (r = -.81, p = .028), indicating lower stimulus discrimination following longer duration of use. Follow-up analysis clarified that this was due to the strong positive correlation of opioid use

duration and SCR to the $CS_{neutral}$ (r = .90, p = .007). However, this finding is based on only 7 participants, which has to be taken account when interpreting it.

6.4 - Discussion

In this chapter, it was my aim to characterise instrumental aversive learning following chronic opioid use. Using a two-part aversive learning paradigm including Pavlovian fear conditioning and subsequent instrumental learning, I compared patients with opioid addiction to patients with chronic opioid use without addiction as well as healthy participants.

Pavlovian fear conditioning was found to be impaired in patients with chronic opioid use. While healthy participants expressed higher physiological responses to tartans paired with aversive compared to neutral stimuli, both patient groups failed to show such stimulus discrimination. These findings match the results of the previous chapter on fear conditioning, where the same pattern was detected. Also, similar to the last chapter, physiological responses to the unconditioned stimuli did not differ between groups, indicating intact initial processing of aversive stimuli. Additionally, groups did not rate the unpleasantness of the US differently; opioid-addicted patients, however, reported less arousal in response to the aversive US compared to the other groups. Intriguingly, Chapter 5's finding of stimulus discrimination being negatively associated with the duration of opioid use in non-addicted but not in opioid-addicted patients could be replicated in the present chapter despite a different sub-sample of patients (only three participants from this group had SCR data for both tasks), the different task environment (behavioural lab vs. MRI scanner), and the different US utilised (unpleasant auditory/visual stimulation vs. electrical shock). This further reinforces the relationship between opioid use and impaired emotional learning in patients treated for chronic pain; nevertheless, from the present data it is not possible to deduct whether the chronic opioid use did indeed cause the fear conditioning deficits.

In accordance with my first hypothesis, patients with chronic opioid use showed lower avoidance accuracy compared to healthy participants, meaning that patients less often selected the tartan likely to avoid money loss. Opioid-addicted patients did not differ from non-addicted patients in avoidance accuracy. These findings imply impaired avoidance learning following chronic opioid use as patients failed to optimally adapt their behaviour to the outcome structure of the task. This could be due to several reasons. Negative reinforcement, the mechanism underlying avoidance learning, states that an action is more likely to be selected if it has led to the omission of an undesirable outcome in the past (Sutton & Barto, 2018). Within this framework, action selection is based on the comparison of the aversive values of multiple competing options, which are acquired over time. Aversive values of actions that have repeatedly led to a negative outcome should increase; aversive values of actions that have

repeatedly led to avoidance of a negative outcome should decrease. In the instrumental learning paradigm presented in this chapter, consistently selecting the option associated with a lower aversive value (i.e. that has less frequently led to money loss) would result in high avoidance accuracy. Therefore, incorrect or insufficient aversive value updating could cause the reduced avoidance accuracy in patients. Previous evidence of reduced activation of the ventral striatum during avoidance learning in patients in OMT supports this line of thinking (Gradin et al., 2014), as this brain region is critical for value-updating based on avoidance PEs (Eldar et al., 2016; Seymour et al., 2007). Additionally, fMRI findings from Chapter 5 demonstrate activation of the caudate, a sub-region of the striatum encoding aversive prediction errors (Delgado et al., 2008), to safe instead of threatening stimuli during fear conditioning, further demonstrating altered processing of aversive learning signals following chronic opioid use. According to the two-factor learning theory (Mowrer, 1951), the idea of performing an action previously paired with an aversive outcome elicits a conditioned fear response (indicative of a high aversive value), which can be avoided by selecting a different action. Consequently, if fear conditioning is impaired, patients fail to acquire a sense of which action to avoid, which would lead to the reduced avoidance accuracy seen in the present data. Alternatively, it is also possible that patients acquire and update aversive values are correctly, but the action with the lower value is not necessarily selected due to a larger component of random choice, as is common in patients with psychiatric disorders (Pike & Robinson, 2022). Comparing behavioural parameters derived from generative computational models could help to disentangle these processes in patients with chronic opioid use, with lower learning rates pointing towards impaired value updating and decreased inverse temperature indicating higher choice randomness compared to healthy participants.

Conditioned avoidance of stimulus-evoked withdrawal has been proposed as a potential mechanism of relapse following abstinence in opioid addiction (Pantazis et al., 2021). My second hypothesis concerned whether conditioned avoidance differed between patients with and without opioid addiction. In the task presented in this chapter, patient groups showed similar avoidance patterns for tartans conditioned with aversive and neutral stimuli, indicative of no group differences in conditioned avoidance. A non-significant interaction between group, time bin, and stimulus type further indicated no stimulus-dependent differential learning between the groups. However, exploratory analysis revealed that instrumental learning over time, operationalised as gains in avoidance accuracy between the first and the last quarter of trials, differed between the groups dependent on stimulus type. Opioid-addicted patients only increased their avoidance accuracy for tartans previously paired with aversive stimuli but not for either neutrally-paired or familiarised tartans. This demonstrates that the pairing of the tartans with aversive stimuli during fear conditioning modulated patients' instrumental avoidance learning, even though physiological conditioned responses did not differ between aversive and neutral tartans. Importantly, initial avoidance accuracy was not higher for trials with aversive tartans

(indicating different aversive values for the tartans following the fear conditioning) but rather patients only updated aversive values in response to punishment and avoidance outcomes on trials on which aversively conditioned tartans had to be avoided. In computational accounts of reinforcement learning, this would correspond to a higher learning rate (i.e. stronger PE-based updating of aversive values; Sutton & Barto, 2018) on aversive compared to other tartans. Non-addicted patients, on the other hand, only demonstrated avoidance learning for tartans previously presented but not part of differential fear conditioning. The finding that opioid-addicted but not non-addicted patients readily learn to avoid stimuli previously associated with aversive events is in accordance with the idea that conditioned avoidance plays an important role in the development and maintenance of opioid addiction (Pantazis et al., 2021), although these findings point toward facilitated conditioned avoidance learning rather than increased conditioned avoidance per se. A possible explanation for this finding could be that opioid-addicted patients assigned higher salience to the stimuli paired with the aversive material and this increased salience was required to enable PE learning (Boehme et al., 2015). However, future research incorporating hypothesis-driven computational modelling and neuroimaging is required to further disentangle potentially different mechanisms contributing to impaired instrumental avoidance learning following chronic opioid use with and without addiction.

In this chapter, aversive instrumental learning was found to be impaired in addicted and non-addicted patients with chronic opioid use, possibly due to impaired updating of aversive outcome values. Additionally, the previous finding of impaired fear conditioning despite intact processing of aversive stimuli per se could be replicated in both patient groups. Exploratory analysis of avoidance learning deficits revealed different patterns of avoidance learning in the patient groups, most prominently that opioid addicted but not non-addicted patients showed facilitated avoidance learning on trials where they had to avoid tartans that were previously paired with aversive pictures and sounds. If replicated this differential effect in conditioned avoidance learning could contribute to why some patients with chronic opioid use develop addiction why others do not.

Chapter 7 - Drug-related and non-drug-related Pavlovian-To-Instrumental Transfer

7.1 - Introduction

Instrumental learning, discussed in the previous chapter, relies on the assumption that the selection of behaviour results from weighing multiple potential outcomes against each other and selecting the one associated with the highest value. While humans without doubt possess the capacity for such rational decision-making, other - not necessarily task-related - processes can introduce motivational biases into rational decision-making. One such process is Pavlovian-To-Instrumental Transfer (PIT). PIT describes the capacity of environmental stimuli that have acquired motivational properties by way of Pavlovian conditioning to affect instrumental behaviour (Holmes et al., 2010). This could entail increased approach behaviour in environments associated with reward (appetitive PIT) or increased avoidance behaviour in environments previously paired with aversive events (aversive PIT). PIT is a highly adaptive mechanism, increasing preparedness to carry out behaviours in known contexts, which can save cognitive resources and lead to faster and more vigorous approach or avoidance behaviour (e.g. Garofalo & Robbins, 2017; Huys et al., 2011; Talmi et al., 2008).

However, when the use of psychoactive drugs is concerned, PIT has been suggested to contribute to the development and maintenance of drug addiction (Everitt & Robbins, 2016). Initial drug use is often perceived as highly pleasurable (i.e. drug use induces pronounced "liking") and, through Pavlovian conditioning, stimuli paired with drug effects (e.g. a person's local bar, the smell of cigarette smoke) can elicit a strong desire to repeat use (high "wanting" or incentive salience; Robinson & Berridge, 1993). PIT incorporates incentive salience into instrumental decision-making and biases action selection towards drug use once these stimuli are re-encountered, even when, based on a "sober" comparison of outcome values, use might be the less optimal behavioural choice. The process of incentive sensitisation, i.e. attributing high incentive salience to drug-related stimuli while incentive salience of not drug-related stimuli decreases, has been suggested to underlie the compulsive nature of drug use that characterises addiction (Robinson & Berridge, 1993), where drug "liking" and "wanting" are progressively decoupled and drug use continues despite the (explicit) desire to stop. Importantly, due to the long-term persistence of Pavlovian associations, incentive sensitisation of drug-related stimuli promoting drug use via PIT can lead to a sustained vulnerability of craving-induced relapse, even following prolonged abstinence (Berridge & Robinson, 2016).

In experimental settings, learning paradigms can be utilised to investigate neural (Geurts et al., 2013; Talmi et al., 2008), pharmacological (Hebart & Gläscher, 2015), and clinical (Garbusow et al., 2022) correlates of appetite and aversive PIT. In these tasks, participants first learn to perform behaviours to obtain rewards or avoid aversive outcomes (instrumental learning phase). Then, a new set of stimuli is

repeatedly paired with either the same (specific PIT) or different (general PIT) outcomes as in the first phase (Pavlovian learning phase). Finally, the instrumental task is reinstated (commonly in extinction) while the Pavlovian stimuli are simultaneously presented (PIT test phase), e.g. as background images. The degree to which appetitive or aversively conditioned background stimuli alter instrumental behaviour in this phase is indicative of PIT magnitude and can be used to compare incentive salience attributed to the background stimuli.

In the last decade, studies have begun to investigate PIT effects in human drug addiction, primarily testing the relationship between incentive salience and prospective relapse. For example, in a study testing recently detoxified patients addicted to alcohol, the extent of PIT-related NAcc activation was predictive of relapse within the following three months, suggesting that attributing higher incentive salience to Pavlovian cues could render individuals vulnerable to maintaining drug addiction (Garbusow et al., 2016). Additionally, prospectively abstaining patients showed stronger reductions in approach behaviour when alcohol-related (but not neutral) stimuli were presented compared to relapsing patients, which could mean that alcohol-cue induced approach inhibition serves as a factor protecting against relapse (Schad et al., 2019).

While no study to date has investigated appetitive or aversive PIT in patients with opioid addiction, a strong case can be made that PIT also plays a role its development and maintenance. Contrasting the long-held notion that incentive salience/"wanting" relies primarily on dopaminergic stimulation (Nutt et al., 2015), acute administration of opioidergic drugs also hyperactivates the NAcc and amplifies incentive salience (S. Peciña & Berridge, 2013), which would imply stronger appetitive PIT elicited by opioid-use related cues. In line with this, pictures of opioid use-related paraphernalia have been demonstrated to induce craving in patients addicted to opioids (e.g. McHugh et al., 2016), underscoring the idea of "wanting" elicited by the presentation of drug-related stimuli. Cue-induced craving is reliably associated with future drug use (Vafaie & Kober, 2022), which could be explained by appetitive PIT. Additionally, evidence from previous chapters of this thesis demonstrated that Pavlovian pairing of neutral stimuli with aversive outcomes facilitated instrumental avoidance learning from these stimuli in opioid-addicted patients (Chapter 6). While the applied task design did not allow to test for PIT, a similar mechanism could underlie these findings: aversive values acquired by way of Pavlovian learning could increase the salience of the stimuli, sharpen prediction-error coding, and thus enable instrumental learning (Boehme et al., 2015). On the other hand, attributing lower incentive salience to opioid-related cues counteracts the development of incentive sensitisation, with a reduced motivational bias towards drug use impedes the transition into compulsive opioid use. Therefore, higher drug-related and lower non-drug-related PIT could increase addiction vulnerability, while lower

drug-related PIT could protect against escalating opioid use, similar to results observed in patients addicted to alcohol (Sekutowicz et al., 2019; Sommer et al., 2020b).

Aims and Hypotheses

In this chapter, my aim is to characterise appetitive and aversive PIT on both approach and avoidance behaviour in patients with chronic opioid use. Specifically, I aim to compare drug-related and non-drug-related PIT magnitudes between patients with opioid addiction and non-addicted patients. For this, both patient groups as well as healthy participants completed two separate versions of a novel PIT paradigm, one of which uses pictures related to the use of prescription opioids or heroin while the other uses neutral pictures in the Pavlovian conditioning phase of the task.

In line with recent research on drug- and non-drug-related PIT in healthy and addicted samples (for a review see Garbusow et al., 2022), I expect all groups to show PIT in both task versions. However, based on the theory that presentation of drug-related stimuli induce disproportionally high and non-drug-related stimuli induce low incentive salience in patients with addiction (Berridge & Robinson, 2016), I hypothesise that the difference between drug- and non-drug-related appetitive PIT is higher in addicted compared to non-addicted patients.

7.2. - Methods

Sample

Out of the 86 participants included in the study, 85 completed both task versions. One patient from the opioid-addicted group aborted the task during the drug condition and reported that they were not comfortable looking at the presented drug-related stimuli. This patient was subsequently excluded from analysis.

Behavioural task

Each participant completed two separate versions of this task on their study visit, at least two hours apart. The versions differed in terms of the conditioned stimuli used in the Pavlovian conditioning phase (see below). In the non-drug-related task version, pictures of differently coloured textile fabrics were used for all three groups. In the drug-related task version, pictures concerning opioid use were used: Opioid-addicted patients were presented with sets of pictures depending on their (preferred) past route of opioid administration, i.e. either prescription pills, paraphernalia related to smoking

heroin or intravenous heroin use; non-addicted patients, who were all using opioids orally, were presented with pictures of prescription pills; and healthy participants, who had never used an opioid drug, were also presented with pictures of prescription pills. How much participants liked the pictures was assessed using a VAS from 0 to 100 (“not at all” to “a lot”) prior to the start of the task. Apart from the conditioned stimuli, the two task versions were identical. Whether participants completed the non-drug-related or drug-related version first was counterbalanced within groups.

PIT was assessed using a novel task similar in structure to previously published paradigms used to assess PIT in clinical samples (for a review see Garbusow et al., 2022). The task consisted of three phases designed to test the transfer of appetite and aversive conditioning on approach and avoidance behaviour. In the first phase (probabilistic instrumental learning), participants were instructed to recommend rental bikes to their customers. Making good recommendations was crucial, as they learned that they would only get payed if customers were satisfied with the recommendation, and lose money otherwise. Therefore, to make as much money as possible, participants had to learn which bikes would satisfy customers, and which would lead to customers demanding their money back. On each trial, participants saw a picture of a bike displayed on the screen for 2 seconds (see Figure 11a). They could choose to either recommend the bike by touching it on the screen (Go response) or refrain by not responding (NoGo response). Afterwards, visual feedback was provided for 1.5 seconds: A picture of a 50ct coin represented monetary gain, while monetary loss was represented by the same coin but crossed out. Over the course of 60 trials, 6 different bikes were shown 10 times each, discriminable by characteristics such as colour or equipment. Three bikes were randomly determined to be “positive bikes”, resulting in monetary reward in 70% and resulting in monetary loss in 30% of responses, while the others were “negative bikes”, resulting in reward in 30% of responses and resulting in loss in 70% of responses. After the last trial, participants received feedback over how much money they earned in total. This part of the task was identical for the drug and the neutral condition, with different sets of bikes used for the two task versions.

In the second phase (Pavlovian conditioning), 5 pictures were presented to participants (CS; either neutral or drug-related, see above), each displayed for 4 seconds on the screen in random order. 2 seconds following the onset of the CS, one of five different monetary outcomes was displayed for the remaining 2 seconds, representing different magnitudes of either monetary reward or loss (US; - €1, - 50ct, ± 0, + 50ct, + €1). Each stimulus-outcome pairing was additionally underscored by a sound played on headphones varying in pitch (100 Hz, 150 Hz, 250 Hz, 350 Hz, 440 Hz; higher pitch indicating higher monetary value). After all five CS-US pairings were displayed once, explicit knowledge of the associations was tested (query trials): Two of the five CS were presented next to each other, with participants asked to select the CS associated with the higher monetary outcome value by touching it

onscreen (see Figure 11b). Afterwards, all five pairings were presented again. This procedure was repeated 18 times.

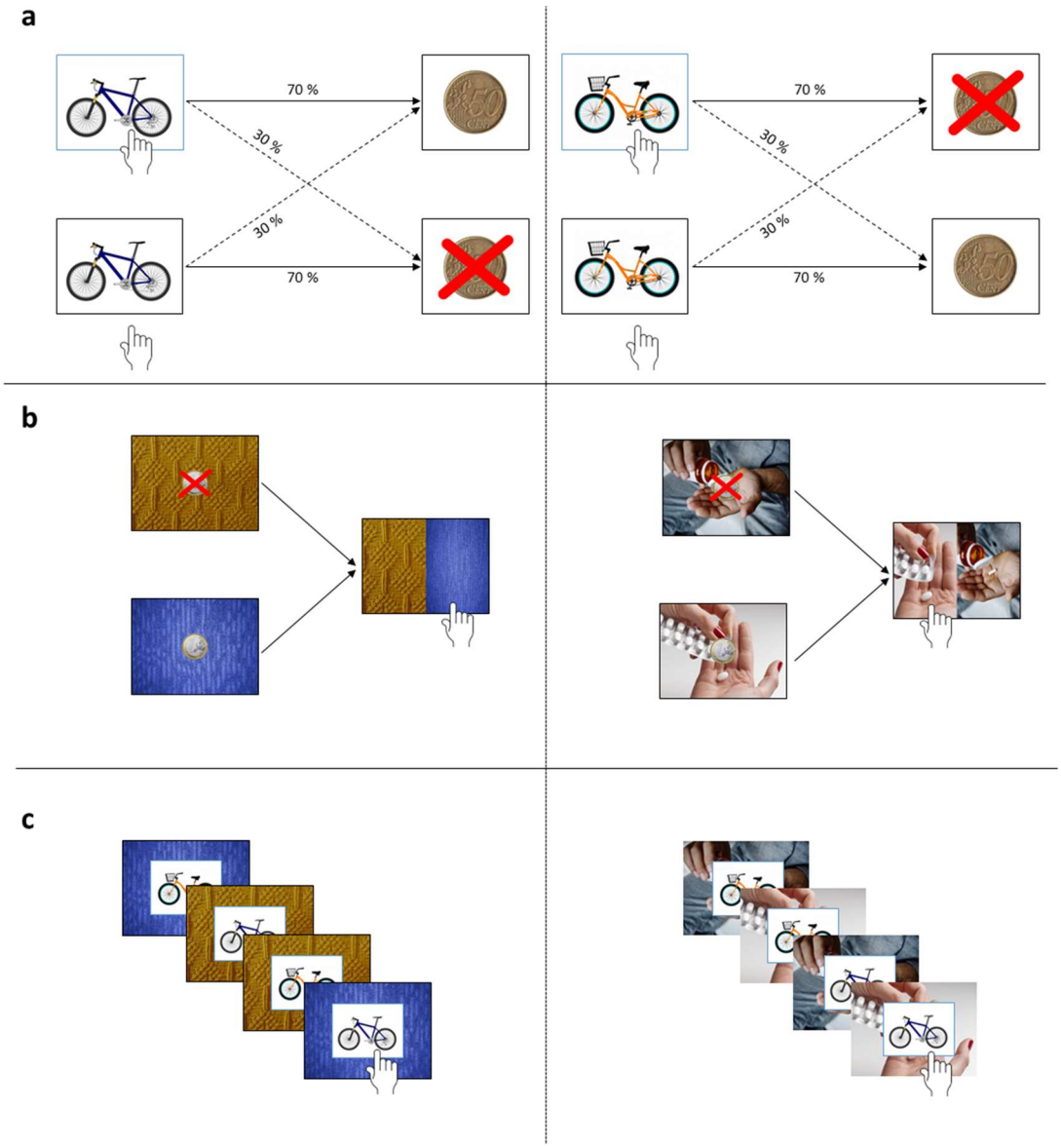


Figure 11. Schematic representation of the PIT paradigm applied in this chapter. Rows represent different task phase: Instrumental learning (a) for approach (left) and avoidance trials (right), Pavlovian conditioning (b), and the PIT test (c), using either non-drug-related (left) or drug-related (right) stimuli.

In the third phase (PIT test), participants were instructed to carry on with the initial task of recommending bikes to their customers in order to gain as much money as possible. This phase was similar to the instrumental learning phase with two important distinctions. First, participants received no feedback on their Go/NoGo responses (the behaviour occurred in extinction) to prohibit further instrumental learning. Second, the CS from the Pavlovian conditioning phase (either non-drug- or drug-related) were displayed in the background of the bikes (see Figure 11c). Each bike was paired with each CS background 7 times, resulting in 35 trials per bike and 210 trials total, with bike and background combined in pseudorandomised order. After the last trial had been completed, participants were

asked to indicate which bike they thought recommending paid off on a VAS from 0 to 100 ("No, it never paid off" – "Yes, it always paid off") to assess explicit knowledge about the reward contingencies. Before and after the task, all patients were asked to indicate how much they would like to use an opioid drug in this very moment on a VAS from 0 to 100 ("not at all" to "very much").

Statistical analysis

Behavioural data from the three task phases and the two task versions were first analysed separately. To test for learning over time in the instrumental learning phase, the 30 approach and 30 avoidance trials were divided into 6 bins of 5 trials each. The percentage of Go responses on "positive bike" trials (approach accuracy) and percentage of NoGo response on "negative bike" trials (avoidance accuracy) were calculated for every participant and every time bin. Group differences in instrumental approach and avoidance accuracy were tested using a 3x6x2 repeated measures ANOVA with accuracy as dependent variable, trial type (approach/avoidance) and time bin (1 to 6) as within-subject factors, and group (healthy participants/opioid-addicted/non-addicted) as between-subjects factor.

For the Pavlovian conditioning phase, the percentage of correct responses on the 18 query trials was compared between groups using a one-way ANOVA with accuracy as dependent variable and group (control/opioid-addicted/non-addicted) as independent variable.

For the PIT test phase, only participants with an accuracy > 50% on the Pavlovian query trials for the respective task version were included into the analyses since PIT can only occur if the CS-US responses are successfully learned in the Pavlovian conditioning phase (Holmes et al., 2010). Response rates (percentage of Go responses) on approach and avoidance trials were computed for each of the five background CS and included into a 3x2x5 repeated measures ANOVA, with response rate as dependent variable, trial type (approach/avoidance) and background (CS_{--}/CS_{-}/CS_0/CS_{+}/CS_{++}) as within-subject factors, and group (healthy participants/opioid-addicted/non-addicted) as between-subjects factor.

Following this, to test whether groups differed in how strongly drug-related and non-drug-related Pavlovian stimuli asserted influence over instrumental behaviour, I compared the (absolute) magnitude of appetitive (CS_{++} - CS_0) and aversive (CS_0 – CS_{--}) PIT on approach and avoidance trials for both task versions, using a repeated measures ANOVA with PIT magnitude as dependent variable, task version (drug-related/non-drug-related), PIT type (appetitive/aversive), and trial type (approach/ avoidance) as within-subject factors, and group (healthy participants/opioid-addicted/non-addicted) as between-subjects factor. For hypothesis testing, I defined planned contrasts for the group-by-task version interaction, testing for higher drug-related PIT appetitive PIT in the opioid-addicted compared to the non-addicted patients.

7.3 - Results

Both patient groups reported similarly low craving pre- and post-task (opioid-addicted pre-task: 15.6 ± 24.0; post-task: 11.3 ± 19.3; non-addicted pre-task: 17.6 ± 25.2; post-task: 11.1 ± 18.9; main effect of time point: $F_{1,45}$ = 3.7, p = .061), demonstrating that the task did not induce substantial craving in patients. Within the group of opioid-addicted patients, 12 (55%) participants used pictures displaying intravenous heroin use, 7 (32%) participants used pictures displaying people smoking heroin, and 3 (14%) participants used pictures showing prescription opioid use as conditioned stimuli, with no differences in craving ratings between the three subgroups (pre-test craving: $F_{2,19}$ = 0.5, p = .611; post-test craving: $F_{2,19}$ = 0.2, p = .823).

Before the start of the task, all participants rated how much they liked to pictures that were later used as Pavlovian stimuli. A main effect of group ($F_{2,82}$ = 3.2, p = .044, η_p^2 = .07) demonstrated that, irrespective of task version, opioid-addicted patients reported lower liking compared to the other two groups. A significant main effect of task version also showed that participants, irrespective of group, rated drug-related pictures lower than non-drug-related pictures ($F_{1,82}$ = 19.4, p < .001, η_p^2 = .19). The group-by-task interaction was not significant ($F_{1,82}$ = 1.1, p = .351).

Non-drug-related task version

In the instrumental learning phase, repeated measures ANOVA demonstrated significant main effects of bin ($F_{5,405}$ = 14.1, p < .001, η_p^2 = .15), trial type ($F_{1,81}$ = 55.0, p < .001, η_p^2 = .40), as well as a significant bin-by-trial type interaction ($F_{5,405}$ = 5.5, $p_{corr.}$ < .001, η_p^2 = .06), but no significant main effect of group ($F_{2,81}$ = 1.8, p = .170), and no interaction of group with trial type ($F_{2,81}$ = 1.2, p = .308) or bin ($F_{10,405}$ = 0.8, p = .602). These results indicate no group differences in approach or avoidance learning (see Figure 12a). Generally, accuracy was higher on approach compared to avoidance trials, which, since relative accuracy gains over time did not differ between the trial types, can be explained by a higher initial accuracy on approach (mean accuracy bin1: 60.5%, CI_{95}: 54.4% - 66.6%) compared to avoidance trials (mean accuracy bin1: 43.1%, CI_{95}: 37.3% - 48.9%; t_{81} = 3.4, $p_{corr.}$ < .001, d = 0.65). Critically, accuracy during the last trials was significantly higher than the chance level of 50% for both trial types (mean approach accuracy bin6: 78.3%, CI_{95}: 73.3% - 83.4%; mean avoidance accuracy bin6: 58.3%, CI_{95}: 52.0% - 64.6%), suggesting successful approach and avoidance learning across groups.

In the Pavlovian learning phase, group differences in accuracy ratings on query trials just failed to reach statistical significance ($F_{2,82}$ = 2.7, p = .077) and showed a trend towards lower accuracy in both patient groups compared to healthy participants (see Figure 12b). Accuracy of three participants (all from the non-addicted group) was below 50%, who were subsequently excluded from the PIT analyses.

In the PIT test phase, results revealed a significant main effect of trial type ($F_{1,79}$ = 73.3, p < .001, η_p^2 = .48), indicating higher response rates on approach compared to avoidance trials (t_{79} = 8.6, $p_{corr.}$ < .001, d = 1.23), and of background ($F_{4,316}$ = 17.2, $p_{corr.}$ < .001, η_p^2 = .18), indicating higher response rates on appetitive background CS compared to neutral background CS (appetitive PIT; CS_{++} vs. CS_0: t_{79} = 3.0, $p_{corr.}$ = .014, d = 0.30) and lower response rates on aversive background CS compared to neutral background CS (aversive PIT; CS_0 vs. CS_{--}: t_{79} = 4.0, $p_{corr.}$ < .001, d = 0.40). The interaction between group and background did not survive Greenhouse-Geisser correction for the violated assumption of sphericity ($F_{8,316}$ = 2.5, $p_{corr.}$ = .063; see Figure 13).

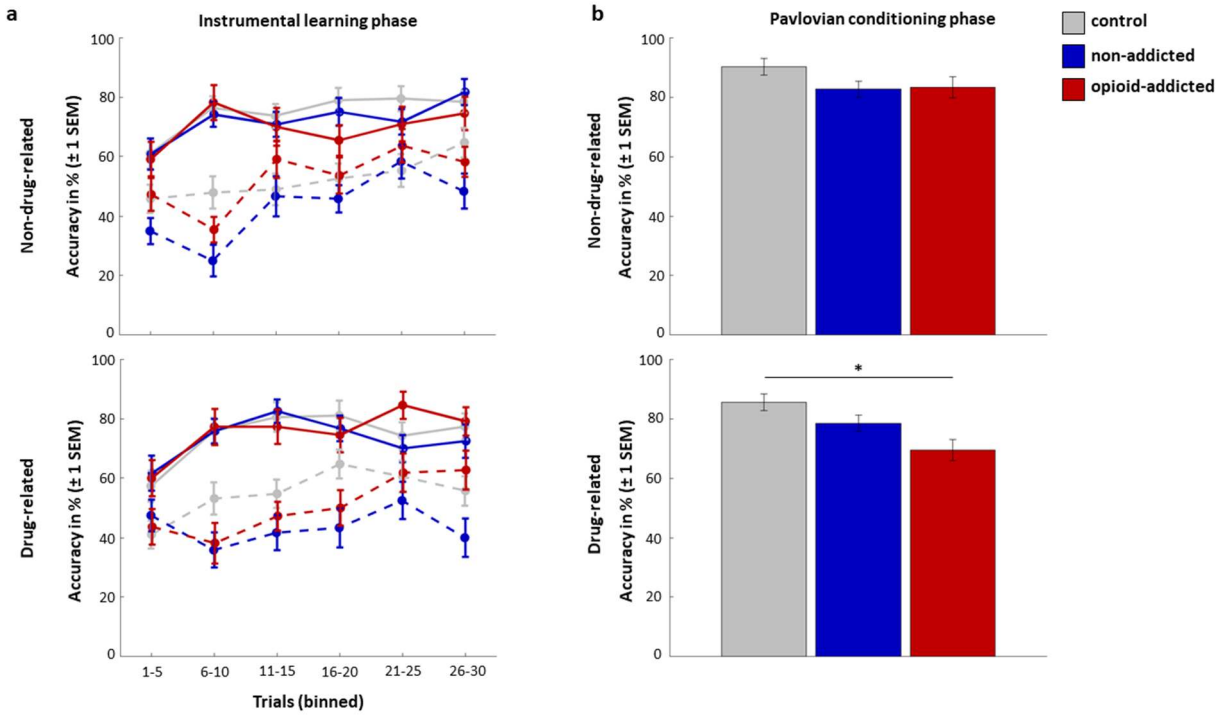


Figure 12. Results of the instrumental (a) and Pavlovian (b) learning phases. Top row refers to the non-drug-related task version, bottom row to the drug-related version. In the instrumental learning results, solid lines represent accuracy on approach trials, dashed lines represent accuracy on avoidance trials.

Drug-related task version

Similar to the non-drug-related task version, repeated measures ANOVA of the instrumental learning phase demonstrated significant main effects of bin ($F_{5,405}$ = 10.5, $p_{corr.}$ < .001, η_p^2 = .12), trial type ($F_{1,81}$ = 52.7, p < .001, η_p^2 = .40), as well as a significant bin-by-trial type interaction ($F_{5,405}$ = 3.8, $p_{corr.}$ = .002, η_p^2 = .04). While the main effect of group was not significant ($F_{2,81}$ = 2.0, p = .142; see Figure 12a), a significant group-by-bin interaction ($F_{10,405}$ = 2.5, $p_{corr.}$ = .006, η_p^2 = .06) revealed lower accuracy in non-addicted patients compared to healthy participants in the fourth bin (trials 16-20; t_{81} = 2.6, $p_{corr.}$ = .036, d = 0.45). As was the case in the non-drug-related task version, initial accuracy (t_{81} = 2.8, $p_{corr.}$ = .006, d = 0.55) as well as accuracy throughout the task (t_{81} = 7.3, $p_{corr.}$ < .001, d = 0.86) was higher

on approach compared to avoidance trials. Accuracy during the last trials was significantly higher than chance on approach trials for all three groups (healthy participants: 77.4%, CI_{95}: 68.4 - 86.3%; non-addicted patients: 79.1%, CI_{95}: 60.9 - 84.1%; opioid-addicted patients: 72.5%, CI_{95}: 69.0 - 89.1%), but for neither group on avoidance trials (healthy participants: 55.8%, CI_{95}: 45.7 - 76.5%; non-addicted patients: 40.0%, CI_{95}: 26.6 - 53.4%; opioid-addicted patients: 62.7%, CI_{95}: 49.0 – 76.5%).

In the Pavlovian learning phase, group differed significantly in accuracy ratings on query trials ($F_{2,82}$ = 7.1, p = .001, η_p^2 = .15; see Figure 12b). Post-hoc tests revealed lower accuracy in opioid-addicted patients compared to healthy participants (t_{82} = 3.8, $p_{corr.}$ < .001, d = 0.98). In total, six participants (one healthy participant, four opioid-addicted patients, and one non-addicted patient) showed choice accuracy under 50% and were subsequently excluded from the PIT analyses.

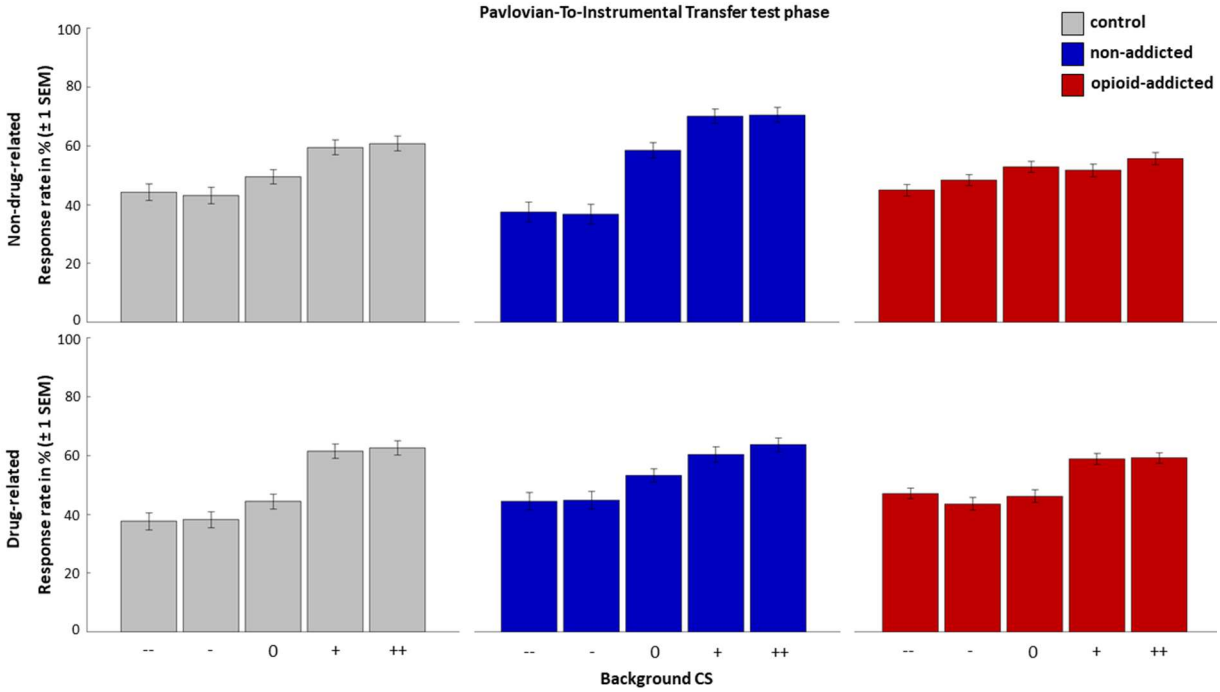


Figure 13. Results of the PIT test phase. Top row refers to the non-drug-related task version, bottom row to the drug-related version.

In the PIT test, repeated measures ANOVA revealed a significant main effect of trial type ($F_{1,76}$ = 33.8, p < .001, η_p^2 = .31) and of background ($F_{4,304}$ = 18.8, $p_{corr.}$ < .001, η_p^2 = .20), indicating higher response rates on approach compared to avoidance trials (t_{76} = 5.8, $p_{corr.}$ < .001, d = 0.9) as well as higher response rates on appetitive background CS compared to neutral background CS (CS_{++} vs. CS_0: t_{76} = 4.1, $p_{corr.}$ < .001, d = 0.4) and lower response rates on aversive background CS compared to neutral background CS (CS_0 vs. CS_{--}: t_{76} = 2.5, $p_{corr.}$ = .047, d = 0.2). The group-by-background interaction was not significant ($F_{8,304}$ = 0.8, $p_{corr.}$ = .477), indicating comparable PIT effects between patients and healthy participants (see Figure 13). There was a trend towards a significant group-by-trial type interaction

($F_{2,76}$ = 2.5, $p_{corr.}$ = .086), suggesting higher response rates on approach compared to avoidance trials in healthy participants (t_{76} = 5.6, $p_{corr.}$ < .001, d = 1.3) and opioid-addicted patients (t_{76} = 5.6, $p_{corr.}$ < .001, d = 1.1) but not in non-addicted patients (t_{76} = 1.7, $p_{corr.}$ = .103).

Comparisons between drug-related and non-drug-related PIT

To test whether groups differed in how strongly neutral and drug-related Pavlovian stimuli asserted influence over instrumental behaviour, I compared the (absolute) magnitude of appetitive PIT (CS_{++} - CS_0) and aversive PIT (CS_0 – CS_{--}) effect on approach and avoidance behaviour for both task versions. Subsequent repeated measures ANOVA revealed a significant main effect of trial type ($F_{1,73}$ = 7.1, p = .010, η_p^2 = .09), with post-hoc tests indicating a slightly stronger PIT effect on approach compared to avoidance trials across groups (t_{73} = 2.7, $p_{corr.}$ = .010, d = 0.05). Furthermore, there was a trend towards a significant PIT type-by-task version interaction ($F_{1,73}$ = 3.7, p = .059), suggesting higher appetitive but lower aversive PIT in the drug compared to the non-drug-related task version. While the group-by-task version interaction did not reach significance ($F_{2,73}$ = 2.4, p = .100), more fine grained planned contrast analysis revealed a significantly higher difference between drug-related and non-drug-related appetitive PIT on avoidance trials (t_{73} = 2.0, p = .049, d = 0.7), but did not on approach trials (t_{73} = 1.5, p = .137), in the opioid-addicted compared to the non-addicted patients (see Figure 14).

When asked for explicit contingency knowledge after the end of the task, participants correctly reported that selecting positive bikes rather than negative bikes resulted in the better outcome ($F_{1,82}$ = 70.9, p < .001, η_p^2 = .46). A trend towards a significant group-by-stimulus interaction ($F_{2,82}$ = 2.9, p = .063) suggested somewhat higher ratings for the negative bikes in the non-addicted patients compared to the two other groups. Task version had no influence on the ratings ($F_{1,82}$ = 0.1, p = .761).

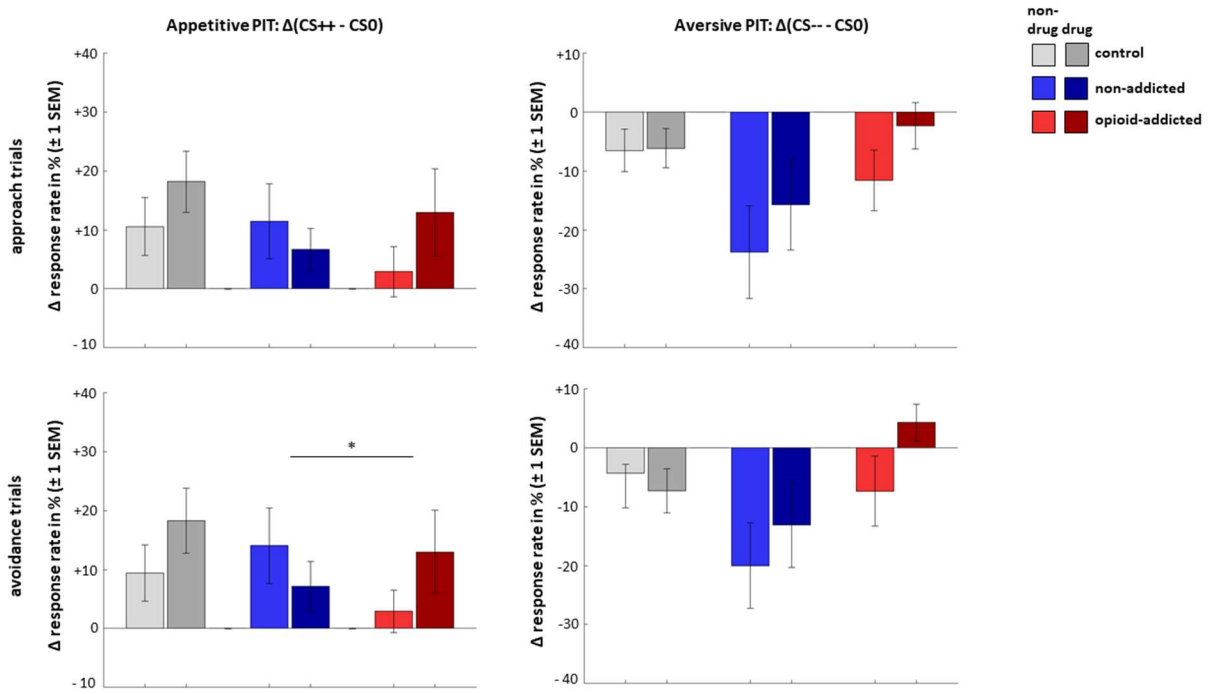


Figure 14. Differences in PIT magnitudes between the two task versions. Left column refers to appetitive PIT, right column to aversive PIT; top row refers to PIT effects on approach behaviour, bottom row to PIT effects on avoidance behaviour.

Associations with opioid use characteristics

In the group of opioid-addicted patients, no significant correlations between opioid use characteristics and behavioural measures were found. However, craving before and after the task was significantly and positively correlated with current daily opioid dose (pre: $r = 0.45$, $p = .034$; post: $r = 0.45$, $p = .036$) and negatively correlated with duration of opioid use (pre: $r = -0.47$, $p = .029$; post: $r = -0.51$, $p = .015$). Additionally, pre-task craving was positively correlated with self-reported obsessive-compulsive drug use (OCDUS score; $r = 0.77$, $p < .001$). In the group of non-addicted patients, current daily opioid dose was negatively associated with the magnitude of the aversive PIT effect on avoidance trials in the non-drug-related task version ($r = -0.53$, $p = .011$). Age of first opioid use was negatively correlated with the magnitude of the appetitive PIT effect on avoidance trials in the drug-related task version ($r = -0.43$, $p = .037$).

7.4 - Discussion

The aim of this chapter was to characterise drug-related and non-drug-related PIT in patients with chronic opioid use. Therefore, patients with and without opioid addiction as well as healthy participants completed two separate versions of a three-phase PIT paradigm, one of which used opioid-related, the other non-opioid-related pictures as Pavlovian CS.

As PIT describes an interplay between instrumental action-outcome and Pavlovian stimulus-outcome associations, adequate acquisition of both is critical (Holmes et al., 2010), which is why it is important to first validate learning success before discussing the hypothesis testing pertaining to PIT. Firstly, instrumental learning did not differ between the groups in the non-drug-related task version. All three groups demonstrated intact reward and avoidance learning, although avoidance accuracy was significantly lower compared to approach accuracy. This difference was already apparent in early trials, when participants largely defaulted towards Go responses on both trial types. While learning that Go responses result in rewards is rather intuitive, avoidance learning from negative feedback to Go responses is considerably less effective (Guitart-Masip et al., 2011, 2012), likely due to a neural coupling of valence and vigour linking reward to action and aversion to inhibition. Importantly, in the last bin of trials, all three groups showed approach and avoidance accuracies above chance level, demonstrating successful learning across groups. In the drug-related task version a slightly different pattern emerged. While reward learning was equally successfully in the three groups, participants did not learn to withhold responses on trials with a high probability of monetary loss, indicated by avoidance choice accuracies indifferent from chance level. Accuracy was lowest in non-addicted patients, which agrees with the instrumental avoidance learning deficits in this group reported in the previous chapter; however, opioid-addicted patients showed similar avoidance learning to healthy participants, opposing earlier results. While the two task designs differed in some relevant features which might influence results (active vs. passive avoidance learning; single-lever Go/NoGo design vs. two-lever forced choice design), the fact that healthy participants also performed at chance level suggests that the task had too few trials to enable successful avoidance learning (30 trials compared to 120 trials in the task utilised in Chapter 6), which might cover up potential group differences, similar to the results in Gradin et al. (2014) discussed previously. The fact that avoidance learning was successful in the non-drug-related task version but not in the drug-related version is puzzling as instrumental learning did not differ between the two task versions. However, since average avoidance accuracy was merely around 5 percentage points higher in the non-drug-related compared to the drug-related version, this might represent statistical variance rather than being rooted in the task design.

Secondly, for PIT to manifest, participants have to successfully acquire CS-US associations during the Pavlovian learning phase. This was confirmed by asking participants to select the CS that had been paired with the higher-outcome US. While the majority of participants learned the correct pairings, three and six participants respectively had to be excluded from subsequent PIT analysis as their accuracy on Pavlovian query trials was below 50 percent. This cut-off was selected based on previous studies comparing PIT in healthy participants and addicted patients (Garbusow et al., 2016); it could be argued that it is too low to ensure that only patients with robust contingency knowledge are included into PIT analysis. In the non-drug-related task version, a statistical trend suggested lower

accuracy in both patient groups compared to healthy participants. This matches the impaired Pavlovian fear conditioning demonstrated previously in this thesis (see Chapters 5 and 6) and potentially extends them to include deficits in appetitive Pavlovian conditioning in patients with chronic opioid use. While endogenous opioids are critical for aversion processing, they also play a key role in reward processing (Fields & Margolis, 2015), and an altered opioid systems due to chronic use might therefore also impact the association of monetary gains with previously neutral stimuli. However, as the present task design does not allow to differentiate between appetitive and aversion Pavlovian conditioning, this is speculative and has to be formerly investigated in future research. The finding that opioid-addicted patients showed lower accuracy compared to non-addicted patients in the drug-related task version provides first evidence of differential processing of drug-related stimuli between the two groups. Opioid-addicted patients have been reported to show increased physiological and neural responses to drug-related stimuli (e.g. Back et al., 2014; Dejoie et al., 2024) and such stimuli often induce an attentional bias in patients (MacLean et al., 2018), which was found to be notably smaller in non-addicted opioid users (Garland et al., 2013). This could suggest that the presentation of opioid-related cues distracted addicted patients by moving attention away from the task at hand, which then might result in lower task accuracy, indicating different affective properties associated with drug-related images in the two groups.

Despite reduced avoidance learning in non-addicted patients and selective impairments in Pavlovian conditioning in addicted patients, I found no group difference in PIT in both task versions, with background stimuli conditioned with money loss reducing and stimuli conditioned with money gain increasing response rate, in accordance with previous PIT studies (Geurts et al., 2013; Huys et al., 2011). The magnitude of appetitive and aversive PIT were comparable, which agrees with most previous studies (Geurts et al., 2013; Talmi et al., 2008; but see Garbusow et al., 2016; Hebart & Gläscher, 2015).

Based on the assumption that addicted patients attribute higher incentive salience to drug-related and lower salience to non-drug-related stimuli (Robinson & Berridge, 1993), I tested the hypothesis of a larger difference between drug- and non-drug-related appetitive PIT in addicted compared to non-addicted patients. The results reported in this chapter are in partial agreement with this hypothesis: The difference in appetitive PIT magnitude between the drug- and non-drug-related task versions was higher for opioid-addicted patients compared to non-addicted patients, but only on avoidance and not on approach trials. This is to say that when value-based decision-making suggests inhibiting Go responses in conflict with reward-associated Pavlovian background stimuli, drug-related images bias behaviour more towards responding than non-drug-related images in addicted patients, but not in non-addicted patients. Following the idea of incentive sensitisation (Robinson & Berridge, 1993), it is

possible that higher incentive salience attributed to drug-related stimuli potentiates the rewarding properties of the Pavlovian background stimulus and therefore promotes a Go response in opioid-addicted patients, potentially due to additive contributions to a mesocorticolimbic reactivity state in the moment of action selection (Berridge, 2012). This would, however, not explain why a similar effect was absent on approach trials. Alternatively, since instrumental avoidance learning was less successful compared to approach learning, it could have been more readily influenced by Pavlovian stimuli, as animal studies show that shorter instrumental training is predictive of higher PIT magnitudes (Holmes et al., 2010). In any case, these results demonstrate a key difference between opioid-addicted and non-addicted patients: drug-related cues that have been paired with positive events in the past can contribute to override value-based NoGo responses. This provides a valuable model for explaining why some individuals relapse after re-exposure to drug-related environmental cues while others maintain abstinence: When faced with the decision to use or abstain, the presence of drug-related environmental cues biases behaviour towards approach behaviour, even though opioid use results in negative consequences and should be avoided.

In this chapter, multiple important and novel findings have come to light. While not the primary focus of this chapter, instrumental reward learning in patients with chronic opioid use was found to be comparable to healthy controls, in agreement with previous research (Gradin et al., 2014; Myers et al., 2017). This contrasts findings of impaired instrumental avoidance learning in patients with chronic opioid use demonstrated in previous chapters, proving that instrumental learning is not per se altered following chronic opioid use but deficits are specific to learning from aversive information. In addition, while Pavlovian learning appears to be generally reduced in patients with chronic opioid use, opioid-addicted patients perform worse than non-addicted patients in pairing monetary outcomes to drug-related compared to non-drug-related stimuli, which could reflect a difference in attentional processing of these stimuli. This presents a possible vulnerability to opioid addiction, as a higher attentional bias towards opioid-related stimuli has been demonstrated to predict a higher likelihood for relapse in opioid addiction (Marissen et al., 2006) as well as for opioid misuse in chronic pain patients (Garland & Howard, 2014). Higher attentional bias to drug-related cues has also previously been linked to higher incentive salience in addicted patients (Vollstädt-Klein et al., 2012), which fits with the finding of higher appetitive PIT (reflective of high “wanting”) towards drug-related images in opioid-addicted patients. In accordance with the notion that high incentive salience towards drug-related stimuli contributes to compulsive drug use (Robinson & Berridge, 1993), motivational properties that drug-related stimuli acquire through Pavlovian conditioning can elicit increased approach behaviour even though inhibiting behavioural responses would result in better outcomes. The fact that opioid-addicted patients showed this effect while non-addicted patients did not suggests

that this could be mechanism driving the development of addiction and potentially explain why one group has developed addiction while the other did not.

Chapter 8 - Avoidance habits in contrast to appetitive habits

8.1 - Introduction

Instrumental avoidance learning, outlined in the previous chapters, is based on acquiring knowledge about causal action-outcome contingencies and selecting the action associated with the least aversive outcome. However, comparing behavioural options based on their expected outcome value (i.e. goal-directed behaviour) requires time and cognitive resources, and can often be unnecessary in routine settings. Everyday experience teaches us that following extensive repetition of a behaviour, representation of the associated outcome is no longer required. Instead, behaviour is carried out automatically when encountering familiar situations (e.g. running on "autopilot" during our morning commute from home to workplace): The initial goal-directed behaviour is neglected in favour of an automatic stimulus-response habit (Robbins & Costa, 2017). Adaptive behavioural control requires a context-dependent balance of habits and goal-directed behaviour: Habits can be executed quickly, come at low computational costs, and can be safely performed in familiar environments; goal-directed behaviour is flexibly adjustable to environmental changes, should guide behaviour in novel environments, and override habitual responding in high-stake situations.

An imbalance between the two can, however, lead to maladaptive behavioural patterns as an overreliance on habits at the expense of goal-directed behaviour has been proposed to lie at the root of drug addiction (Everitt & Robbins, 2005, 2016). This theory states that initial drug use is goal-directed, motivated by its expected rewarding outcome. If drug use repeatedly leads to a reward, the action-outcome contingency characterising goal-directed behaviour is reinforced, increasing probability of repeated use in the future. Simultaneously, repeated use results in the progressive formation of outcome-independent stimulus-response associations (habit learning) and drug use gradually becomes more and more habitual. Following continued use, negative consequences of drug use begin to replace the initial rewarding effects, creating a conflict between goal-directed action (promoting to stop drug use as the associated outcome is now aversive) and outcome-insensitive habitual behaviour, which, triggered by drug-related environmental cues, continues to promote use. Depending on the relative strength of learned goal-directed and habitual associations, individuals with either impaired goal-directed learning or increased habit learning are biased towards automatic, habitual decision-making increasing the likelihood of stimulus-triggered drug use. When an individual fails to exert regulatory control over habitual drug use, i.e. override the habitual urge by exerting top-down control, use becomes compulsive, a central defining characteristic of drug addiction (Everitt & Robbins, 2016). The habit theory of addiction was based on evidence from experimental rodent research, where the role of over-training (Belin & Everitt, 2008), pharmacological drug effects (Dickinson et al., 2002; F. J. Miles et al., 2003) and inter-individual vulnerability (Deroche-Gamonet et

al., 2004) on habit formation and the gradual transition into compulsive drug use were established. More recently, translation of experiments into human research has allowed to test the balance between goal-directed and habitual behaviour in patients with drug addiction, manipulating different aspects of goal-directedness (De Wit & Dickinson, 2009). According to Heyes & Dickinson (1990), a behaviour can be considered goal-directed when it is performed because it is believed to result in a desirable outcome. Is either the desire criterion or the belief criterion not met (i.e. the goal is undesirable or the behaviour is not believed to cause the outcome), there is no goal-directed reason to carry out the behaviour. If carried out regardless, the behaviour can be considered a habit.

Outcome devaluation, one validated paradigm to test the balance between goal-directed and habitual behaviour, manipulates the desirability of the outcome: After a prolonged training in which participants learn that responding to certain stimuli leads to an outcome associated with reward, they are informed that some outcomes are now dissociated from reward (devalued) while others remain linked to obtaining a reward (valued). Stopping responses to the stimulus associated with the devalued outcome indicates goal-directed behaviour, while continued responding violates the desire criterion and indicates outcome-insensitive habitual behaviour (De Wit et al., 2009). In such a task, patients addicted to cocaine failed to reduce responding to the stimulus associated with the devalued outcome (Ersche et al., 2016), demonstrating a bias towards habitual behaviour. Notably, this behavioural effect could have several causes: In addicted patients, goal-directed learning could be impaired (resulting in weaker action-outcome associations), habit learning could be potentiated (resulting in stronger stimulus-response associations) or regulatory control over behaviour arbitrating between the two could be insufficient (Vandaele & Janak, 2018). Neuroimaging can help to differentiate between the mechanisms, as distinct neural networks have been identified as neural substrates of goal-directed (caudate, anterior putamen, vmPFC; De Wit et al., 2009, 2012a) and habit learning (posterior putamen, pre-motor cortex; De Wit et al., 2012a; Dolan & Dayan, 2013). Intriguingly, one study tested behavioural responses to outcome devaluation in patients addicted to alcohol using fMRI (Sjoerds et al., 2013). Here, a higher reliance on stimulus-response habits in patients was complemented by higher activity in the posterior putamen, indicating increased habits rather than decreased goal-directedness to underlie the behavioural effect.

Contingency degradation is another experimental approach to test the balance between goal-directed and habitual behaviour. Here, the belief criterion of goal-directedness is targeted: After participants had learned that responding to a stimulus caused a rewarding outcome, non-responding progressively also delivered the reward, uncoupling responses from reward, and rendering the outcome action-independent (Vaghi et al., 2019). In this task, patients addicted to cocaine reduced responding to a significantly lower degree compared to healthy participants after the action-outcome associations had

been degraded (Ersche et al., 2021). Interestingly, patients still believed that responding caused the reward, suggesting that behaviour turned more habitual because of a lack of awareness concerning the updated action-outcome contingencies. Furthermore, the degree to which patients adapted their behaviour in response to altered contingencies was negatively associated with the duration of cocaine use, underscoring the notion of a progressive shift into from goal-directed into habitual drug use.

The use of opioids, in contrast to other psychoactive drugs such as cocaine (Badiani et al., 2011), is not driven by the desire to obtain a reward but rather motivated by avoiding aversive states (Koob, 2020). Therefore, the development of opioid addiction can be conceptualised as escalating avoidance behaviour, initially goal-directed and then gradually shifting into habitual avoidance, which can lead to compulsive use if regulatory control over avoidance behaviour is insufficient. The distinction between appetitive (i.e. reward-based) and avoidance habits is crucial: In their aforementioned study, Ersche et al. (2016) found increased reliance on appetitive habits in patients addicted to cocaine but observed no such tendencies when testing avoidance habits, demonstrating valence-specific habit formation in this patient group. The authors suggested that, given the importance of avoiding withdrawal, patients addicted to opioids might show an opposite pattern. While the centrality of negative reinforcement for the development and maintenance of opioid addiction has been stressed repeatedly (Koob, 2020; Pantazis et al., 2021), avoidance habits have not been investigated to date. Comparing avoidance habits in patients addicted to opioids to patients who also use opioids but never developed addiction could be particularly fruitful, potentially shedding light on why some chronic opioid users transition from goal-directed use into addiction while others do not.

To summarise, gradually shifting from goal-directed to habitual behaviour is an adaptive process, saving time and cognitive resources. However, the ability to exert goal-directed control over habits in high-stakes situations is critical. Compulsive drug use, a hallmark criterion for drug addiction, has been proposed to be caused by an over-reliance on stimulus-response habits that are detached from an individual's goals and cannot by inhibited despite adverse consequences. Given the role of avoidance in opioid use, increased habitual avoidance could underlie the development of opioid addiction, in which case it would be observable in opioid-addicted patients but not in chronic users that never developed opioid addiction.

Aims & Hypotheses

In this chapter, I aim to investigate goal-directed and habitual avoidance behaviour in opioid-addicted and non-addicted patients. To detect a potential imbalance between the two systems, I will apply two separate behavioural paradigms, outcome devaluation and contingency degradation, targeted at the

desire and the belief criteria of goal-directed behaviour. To test the specificity of the proposed over-reliance of habits to the domain of negative reinforcement (avoidance habits), I will also apply an appetitive version of the contingency degradation paradigm. Additionally, I will use fMRI during the outcome devaluation task to further disentangle potential contributors to increased habitual avoidance, testing activation patterns in the neural substrates of goal-directed and habitual behaviour.

I hypothesise an increased tendency towards habitual avoidance in patients addicted to opioids compared to non-addicted patients and healthy participants. This is indicated by a higher avoidance response rate to the stimulus associated with the devalued outcome following overtraining as well as a higher avoidance response rate following contingency degradation. Comparing neural activation in brain regions linked to goal-directed and habitual behaviour can help to further disentangle the interplay between the two behavioural systems. Moreover, I hypothesise that this increased reliance on habits is specific to the aversive domain, indicated by no group differences in behaviour following contingency degradation in a separate, appetitive task version.

8.2 - Methods

Sample

Data from all 86 study participants is included in this chapter. For the first task (avoidance learning and outcome devaluation), data from one non-addicted patient was discarded because a Wasp electrode lost contact to their wrist during the task. Furthermore, SCR data from 12 participants was excluded due to noisy recordings (5 healthy participants, 2 opioid-addicted patients, 5 non-addicted patients); their behavioural and fMRI data was included. The first 7 participants included in the study (all from the control group) were scanned using sub-optimal fMRI parameters, which were later adapted, leading to the exclusion of these participants from neuroimaging analysis; behavioural and SCR data were included. Only 72 (30 healthy participants, 20 opioid-addicted patients, 22 non-addicted patients; 84% of the total sample) participants completed the second task (aversive contingency degradation) because it was added late to the study protocol. Additionally, one patient from the opioid-addicted group aborted the task because she experienced the aversive stimulus as unbearable. All 86 participants completed the third task (appetitive contingency degradation).

Task 1 – Avoidance learning and outcome devaluation

Participants completed this task, developed by Gillan et al. (2014, 2015), inside the MRI (see Chapter 3), where they placed their feet in a foot switch, enabling them to navigate the task by tapping their feet against a plastic plate. They were informed that aversive electric shocks could be delivered to either their left or right wrist and their aim was to avoid the shocks. Before the start of the task, shock intensity was individually calibrated (see Chapter 3) and three fractal images were introduced, serving as visual cues which informed participants about action-outcome contingencies. One fractal was paired with a shock to the left wrist, another one with a shock to the right wrist, and a third one without shock, which could be avoided by pressing the corresponding pedal on the foot switch (see Figure 15). In the first part of the task (training phase), participants learned to avoid electrical shocks (instructed instrumental avoidance learning): On each trial, a fractal was displayed, serving either as a warning signal (announcing that a shock was imminent unless the correct pedal was pressed) or as a safety signal (announcing that no response was required). If participants pressed the wrong pedal, pressed both pedals or pressed late, a shock was delivered to the respective wrist. How much time participants had to respond was individually calibrated before the start of the task (see below). Training consisted of four blocks, each containing 30 trials in pseudorandomised order, which included 20 warning signals (10 left, 10 right) and 10 safety signals, which were interspersed by a white-on-black fixation cross presented for 8 s. After training was completed, one of the shock electrodes (left or right, randomly determined) was physically disconnected from the wrist and participants were informed that they could not receive shocks on this wrist anymore (outcome devaluation), while nothing else had changed and the task remained to avoid shocks. To emphasise the devaluation of shocks to one wrist, shocks were delivered to both electrodes and participants confirmed that they had only experienced a shock on the valued wrist. The ensuing devaluation test consisted of 30 trials (10 presentations of the fractal associated with the valued outcome, 10 of the fractal associated with the devalued outcome, 10 safe) in pseudorandomised order. Importantly, the devaluation test was conducted in extinction (no shock in the case of late or wrong responses), so no further instrumental learning could occur.

Before every block of the training phase as well as before and after the devaluation test, participants were asked to rate the shock expectancy associated with each of the three fractals separately on a VAS from 0 to 100 (“not at all” to “very likely”). Key behavioural outcome measures was the rate of correct avoidance responses during the training phase as well as the number of responses to the fractals associated with the valued and devalued outcomes during the devaluation test. To ensure task comprehension, two practice sessions were implemented before the actual task started. The first was conducted outside of the MRI scanner. Participants sat in front of a computer screen and completed a task resembling the actual task described above with the distinction that the inter-trial-interval was

reduced to 3 s, different visual cues were used, and participants had to push arrow keys instead of the foot pedals. A yellow flash appeared on the screen every time the participants committed a mistake that indicated a shock. During the practice session, all participants initially had 750 ms to respond following the onset of the visual cue. After completion of 30 trials (10 left, 10 right, 10 safe), avoidance accuracy was computed (number of correct responses divided by the number of trials). If accuracy was higher or equal to 80%, practice was considered successful and participants entered the MRI scanner; if not, practice blocks were repeated until participants had successfully completed a block with the minimum accuracy. Each time the practice block was repeated, time to respond increased by 100 ms. To familiarise participants with the foot switch, another block of practice followed inside the MRI. Different fractal images were introduced and a shock was indicated by a yellow flash. Time to respond for this practice was set as the time used to successfully complete the practice outside of the scanner, and the actual task started as soon as participants had a practice response accuracy of at least 80%, again increasing time to respond by 100 ms every time a participant had lower than this accuracy. The final time to respond used when successfully completing the second practice session was used for the actual task.

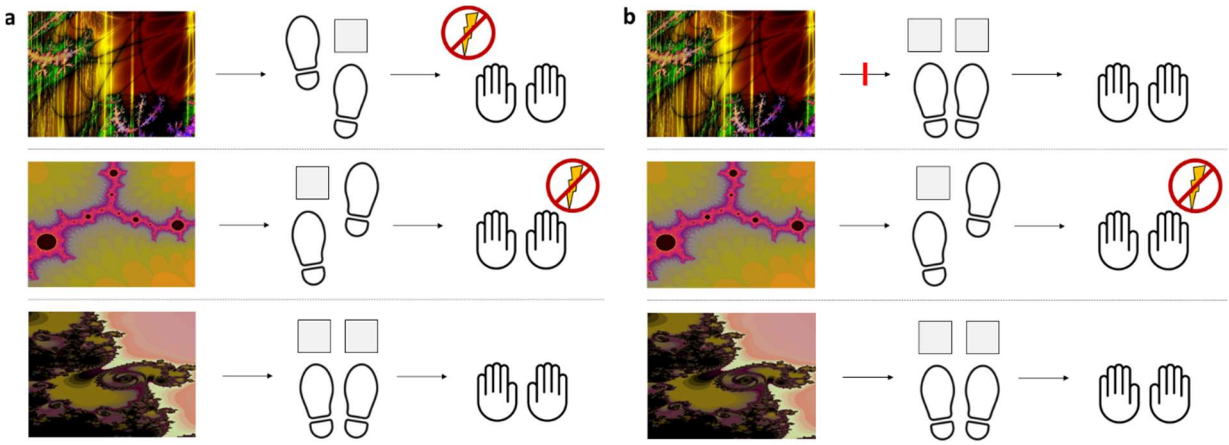


Figure 15. Schematic representation of the avoidance learning and outcome devaluation task. Displayed are action-outcome contingencies from the training (a) and test (b) phase.

Task 2 - Aversive contingency degradation

In this novel task adapted from a previously published appetitive habit task (Ersche et al., 2021), participants were placed in a computerised scenario where they had taken on a job in a nursery, entrusted with taking care of a group of young children. On each trial, participants were presented with the picture of an empty crib while the sound of crying babies was played on headphones (see Figure 16a). Participants were instructed that they could respond to the crib by pressing the space bar, through which they could avoid the crying in some instances, but not in others, while it was also possible to avoid the crying without responding. If the crying was avoided, a picture of a baby smiling

happily appeared inside the crib; otherwise a discontented baby appeared while the crying noise continued. On each trial, the crib was displayed for 1 s, during which a response was possible; feedback (baby picture plus either silence or continued crying) was presented for 2 s. Adding a social evaluation element to the task, participants were instructed that their new colleagues at the nursery would monitor their performance closely and reward them with social media "likes" when the babies were calm and punish them with "dislikes" when the babies were crying. Participants' scores of likes and dislikes were displayed separately on the screen at all times during the task and were continuously updated depending on task performance albeit in a biased fashion aimed at increasing motivation to avoid, with one like added every 200 ms of silence and one dislike added every 140 ms of crying. Participants were instructed that the goal of the task was to minimise the crying, thereby gaining their colleagues respect and successfully completing their probationary period.

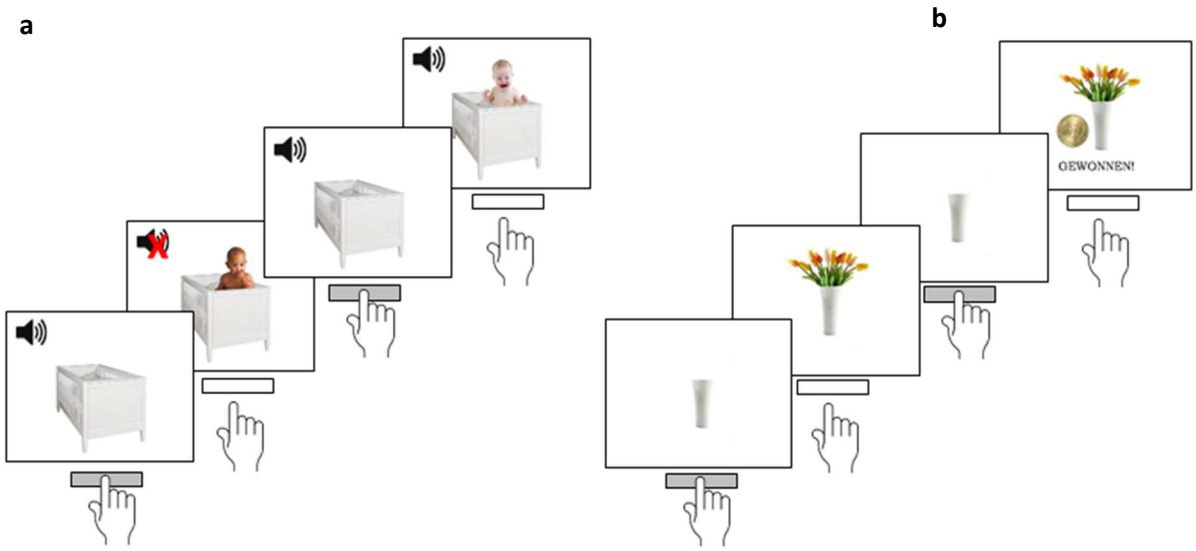


Figure 16. Schematic representation of the aversive (a) and appetite (b) contingency degradation task. Both tasks were completed separately at least three hours apart.

Importantly, while the probability that a button press resulted in avoiding the crying remained fixed at 60% during the course of the task, avoidance probability without response varied throughout the 8 task (100 trials each; see Table 5). During blocks 1, 2 and 3, absence of a response never resulted in avoidance (non-degraded condition). During block 4, non-responding resulted in a 30% chance to avoid the crying (partial degradation condition). During block 5, non-responding resulted in a 60% avoidance probability, matching the avoidance probability for responses and thus rendering the avoidance probability independent from behaviour (full degradation condition). All conditions were repeated in blocks 6 through 8. After the completion of each block, participants were asked to rate their belief that button presses resulted in avoidance (causality judgements) on a VAS from 0 ("entirely unlikely") to 100 percent ("very likely"). After the task had finished, they also rated how unpleasant they found the

crying, on a VAS from 0 (“not particularly unpleasant”) to 100 (“extremely unpleasant”). The primary outcome measure was the mean response percentage in non-degraded, partially degraded or fully degraded blocks as well as the causality judgements following each block.

Table 5. Action-outcome contingencies for appetite and avoidance contingency degradation tasks.

		Task 2 - Avoidance probability		Task 3 - Reward probability	
Block	Condition	Response	No response	Response	No response
1 to 3	no degradation	60%	0%	60%	0%
4	partial degradation	60%	30%	60%	30%
5	full degradation	60%	60%	60%	60%
6	no degradation	60%	0%	60%	0%
7	partial degradation	60%	30%	60%	30%
8	full degradation	60%	60%	60%	60%

Task 3 - Appetitive contingency degradation

Behavioural adaptions to degradation of appetitive contingencies were assessed using a task previously utilised to investigate habitual responding in clinical samples (Ersche et al., 2021; Vaghi et al., 2019). Participants were instructed to respond to the picture of an empty vase by pressing the space bar (see Figure 16b*)*, resulting in a bouquet of flowers as well as in monetary reward in some but not all cases, while it would also be possible to obtain a reward without responding. On each trial, responding was possible for 1 s; visual feedback (bouquet plus potentially image of a 50 ct coin) was presented for 500 ms. The total amount earned in the current block was presented on screen and increased by 50 ct every time a reward was obtained. Participants were instructed to earn as much money as possible.

Similarly to the aversive task version introduced above, the reward probability following a response was 60% during the entire task, while the reward probability without response varied throughout the 8 task blocks (120 trials each; see Table 5). After the completion of each block, participants indicated their belief that responding resulted in reward on a VAS from 0 (“entirely unlikely”) to 100 (“very likely”). Identically to the aversive task version, primary outcome measures were the mean response percentages in non-degraded, partially degraded or fully degraded blocks as well as the causality judgments following each block. To exclude the possibility that differences in motivational value assigned to winning small monetary rewards affect the results, we asked participants beforehand to select how often they would pick up a 50 ct coin if they found it lying on the street on a VAS from 0 (“never”) to 100 (“always”).

Behavioural analysis

Task 1 – Avoidance learning and outcome devaluation

For the training phase, avoidance accuracy was calculated for every block, representing the percentage of correct responses to warning and safe stimuli. Group differences in avoidance accuracy were tested using a 3x4 repeated measures ANOVA with avoidance accuracy as dependent variable, block (1/2/3/4) as within-subjects factor, and group (control/non-addicted/addicted) as between-subjects factor. Planned contrasts compared healthy participants against patients (group contrast weights: 1 -.5 -.5) as well as the patient groups against each other (group contrast weights: 0 -1 1). For the test phase, response rate was computed separately for the stimulus associated with the valued and devalued outcome and compared using a 3x2 repeated measures ANOVA with response rate as dependent variable, stimulus (valued/devalued) as within-subjects factor, and group (control/non-addicted/addicted) as between-subjects factor. Planned contrast were defined for the group-by-stimulus interaction as differences in response rates between the stimulus associated with the valued and devalued outcome in the two patient groups (interaction contrast weights: 0 0 1 -1 -1 1) as well as between patients and healthy participants (interaction contrast weights: 1 -1 -.5 .5 -.5 .5).

Task 2 - Aversive contingency degradation

Avoidance response rates were averaged for each of the three conditions. Group differences were tested using a 3x3 repeated measures ANOVA with response rate as dependent variable, condition (non-degraded/partially degraded/fully degraded) as within-subject factor, and group (control/non-addicted/addicted) as between-subject factor. Planned contrasts compared 1) healthy participants and patients regarding the difference in response rates during non-degraded and fully degraded blocks (interaction contrast weights: .5 0 -.5 -.25 0 .25 -.25 0 .25) as well as 2) the two patient groups against each other (interaction contrast weights: 0 0 0 .5 0 -.5 -.5 0 .5). Causality judgement for the three conditions were analysed in the same manner but without planned comparisons as no a priori hypotheses were specified. A one-way ANOVA was used to test for group differences in the unpleasantness rating.

Task 3 - Appetitive contingency degradation

In similar fashion, group differences in reward response rates were tested using a 3x3 repeated measures ANOVA with response rate as dependent variable, condition (non-degraded/partially degraded/fully degraded) as within-subject factor, and group (control/non-addicted/addicted) as

between-subject factor. Identical to the aversive task version, planned contrasts compared 1) healthy participants with patients and 2) the two patient groups with each other. Causality judgement for the three conditions were analysed in the same manner. A one-way ANOVA was used to test for group differences in how often they would pick up a 50 ct coin.

Physiological recording, pre-processing and analysis

During the course of this task, EDA was recorded and subsequently processed according to Chapter 3. Response window was defined between 500 ms and 4 s following stimulus onset. SCR were analysed separately for the training and the test phase: for the training phase, SCR were averaged by block (block 1 to block 4) and stimulus (warning/safe), for the devaluation test, SCR was averaged by stimulus (valued/devalued/safe). All SCR data reported in this chapter were standardised and normalised (i.e. z-scored and log-transformed). Group differences in physiological responses to the fractal during training were tested using a 3x2x4 repeated measures ANOVA, with SCR as dependent variable, block (1 to 4) and stimulus (warning/safe) as within-subject factor, and group (control/non-addicted/addicted) as between-subject factor. SCR group differences during the devaluation test were tested using 3x3 repeated measures ANOVA, with SCR as dependent variable, stimulus (valued/devalued/safe) as within-subject factor and group (control/non-addicted/addicted) as between-subject factor.

Neuroimaging acquisition, pre-processing, and analysis

For fMRI data acquisition and pre-processing see Chapter 3. Data from the training and test phases were analysed separately. For the training bocks, first level analysis was conducted using individual general linear models for each participant with one regressor for the warning and safe stimulus during every block of the task (8 regressors in total). Stick functions at stimulus onset were convolved with a canonical hemodynamic response function. Additionally, movement parameters with their first and second derivatives were included as nuisance regressors to control for head motion. To restrict analysis to grey matter voxels, two additional nuisance regressors were added including white matter and CSF segmentation maps estimated during pre-processing. For each participant individually, regressor weights were estimated using the pre-processed 4D-images. Beta images of the 8 regressors of interest were then warped into MNI space and smoothed using a 6 mm FWHM Gaussian kernel. For second level analysis, a flexible factorial design was set up, including one regressor for each of the participants (subject factor), one for each of the groups (group factor), and one for each condition (e.g. “Block1_warning”), as well as a total of 24 regressors for the group-by-condition interaction. Statistical

parametric maps for each of the 84 regressors were estimated and analysed using a priori specified ROI analyses, which were conducted by applying small-volume corrections in regions associated with habitual (putamen, preSMA) and goal-directed (caudate, vmPFC/OFC) behaviour smoothed using brain masks included in the WFU PickAtlas Toolbox (version 3.0.5). T-contrasts for flexible factorial designs were specified to test for interaction effects between groups and conditions (Gläscher & Gitelman, 2008). Contrasts were specified to test for group-by-stimulus interactions in the training phase (warning > safe across blocks for either controls > patients or opioid-addicted patients > non-addicted patients) and in the test phase (valued > safe[7] for either controls > patients or opioid-addicted patients > non-addicted patients), respectively. Results were corrected based on family-wise error rates with α-levels set to .05.

8.3 - Results

Task 1 - Avoidance learning and outcome devaluation

Behavioural results

Individually calibrated shock intensities were comparable between the groups ($F_{2,54}$ = 1.7, p = .195). The number of practice blocks required before the start of the task differed between groups (χ_8^2 = 26.2, p < .001), resulting in a higher average time to respond in opioid-addicted and non-addicted patients compared to healthy participants ($F_{2,53}$ = 9.4, p < .001, η_p^2 = .26).

For the avoidance training, repeated measures ANOVA revealed main effects of group ($F_{2,53}$ = 3.4, p = .040, η_p^2 = .12) and block ($F_{3,159}$ = 9.6, $p_{corr.}$ < .001, η_p^2 = .15) as well as a significant group-by-block interaction ($F_{6,159}$ = 2.3, $p_{corr.}$ = .043, η_p^2 = .08). Planned contrasts for the group effect showed that healthy participants avoided more accurately compared to the patients (t_{53} = 2.3, p = .025, d = 0.52), while avoidance accuracy did not differ between the two patient groups (t_{53} = -0.6, p = .572; see Figure 17a). Post-hoc tests on the significant interaction effect demonstrated improvements in avoidance accuracy between blocks 1 and 2 for the healthy participants (+ 8.7%; t_{53} = 4.6, $p_{corr.}$ < .001, d = 0.85) and the opioid-addicted patients (+ 11.4%; t_{53} = 2.7, $p_{corr.}$ = .049, d = 1.12), but not for the non-addicted patients (-0.4%; t_{53} = -0.2, $p_{corr.}$ = 1). Analyses of shock expectancy ratings indicated no differences between the groups (three-way interaction between group, block, and stimulus: $F_{20,530}$ = 0.6, $p_{corr.}$ = .935), demonstrating that all groups were equally successful discriminating between warning and safety signals.

[7] In line with the original article describing fMRI analysis of this task (Gillan et al., 2015), this contrast was selected to investigate the neural signature of habits without potential confounds associated with the devalued stimulus, such as an urge to respond or attempts to supress responding. Selecting this contrast assumes that if avoidance responding has become habitual for the devalued stimulus, responding to the valued stimulus has also become habitual.

For the devaluation test, mean difference in response rates between the stimulus associated with the valued minus devalued stimulus was nominally lower in opioid-addicted patients (11.4% ± 15.7) compared to non-addicted patients (30.7% ± 42.3) or healthy participants (44.3% ± 44.1). However, repeated measures ANOVA indicated a non-significant interaction effect for group-by-stimulus ($F_{2,53}$ = 2.0, p = .140), indicating comparable behavioural responses to the outcome devaluation between groups (see Figure 17b). Planned contrasts revealed no significant differences between the between two patient groups in response rates to the stimuli (t_{53} = 1.0, p = .319), while patients showed a trend towards a smaller differences in response rate to the stimuli compared to healthy participants (t_{53} = 2.0, p = .056). Groups did not differ in their shock expectancy ratings before or after the devaluation test, indicated by a non-significant three-way interaction between group, time point, and stimulus ($F_{4,106}$ = 1.3, $p_{corr.}$ = .277).

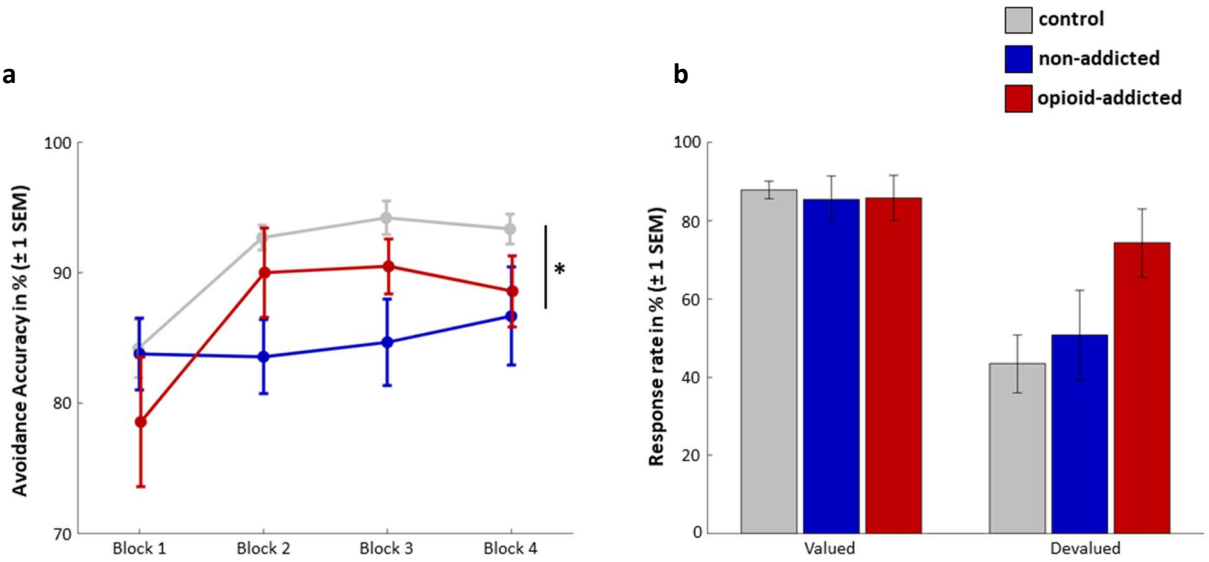


Figure 17. Accuracy during the avoidance training (a) and response rates to the stimuli associated with the valued or devalued outcome in the test phase (b)

Physiological results

Repeated measures ANOVA revealed no group differences for SCR towards warning and safety signals during training (group-by-stimulus interaction: $F_{2,41}$ = 1.9, p = .164). A significant block-by-stimulus interaction ($F_{3,123}$ = 8.6, p < .001, η_p^2 = .18) indicated higher SCR to the warning compared to safe stimuli only in the first block of avoidance training, irrespective of group (group-by-block-by-stimulus interaction: $F_{6,123}$ = 0.8, p = .559). During the test phase, SCR towards the different stimuli did not differ between groups (group-by-stimulus interaction: $F_{4,80}$ = 0.9, $p_{corr.}$ = .451).

Neuroimaging results

ROI analysis for the training phase at $p_{FWE} < .05$ revealed no significant clusters when comparing healthy participants to patients in the Warning > Safe contrast. At $p_{uncorr.} < .001$, ROI analysis showed significantly higher activation in healthy participants in the right vmPFC (xyz = 6 44 -10, $T = 4.8$, $p_{uncorr.} < .001$), right mOFC (xyz = 2 26 -15, $T = 4.0$, $p_{uncorr.} < .001$), left mOFC (xyz = -4 50 -15, $T = 3.3$, $p_{uncorr.} < .001$), right anterior putamen (xyz = 21 10 -9, $T = 3.5$, $p_{uncorr.} < .001$), as well as two separate clusters in the left preSMA (xyz = -8 -3 52, $T = 3.9$, $p_{uncorr.} < .001$; xyz = -2 15 57, $T = 3.6$, $p_{uncorr.} < .001$). Patients showed higher activation in the right posterior putamen (xyz = 28 -10 4, $T = 3.3$, $p_{uncorr.} = .001$; see Figure 18). Exploratory whole-brain analyses conducted at $p_{FWE} < .05$ revealed higher activation in the left dlPFC in patients compared to healthy participants (xyz = -38 22 26, $T = 5.5$, $p_{FWE} = .009$). When comparing opioid-addicted to non-addicted patients using the same contrast, ROI analysis revealed significantly higher activation in the left caudate in opioid-addicted patients (xyz = -16 -8 20, $T = 5.1$, $p_{FWE} = .002$). Exploratory whole-brain analyses identified higher activation in a cerebellar cluster in opioid-addicted patients (xyz = -28 -54 -58, $T = 5.7$, $p_{FWE} = .001$). Non-addicted patients did not show higher activation in any of the defined brain regions at $p_{FWE} < .05$; they did, however, show stronger activation in the right preSMA compared to addicted patients at $p_{uncorr.} < .001$ (xyz = 12 3 57, $T = 3.9$, $p_{uncorr.} < .001$).

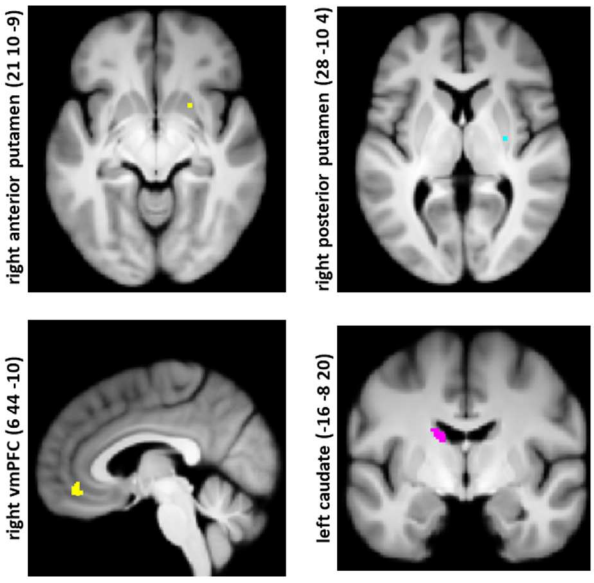


Figure 18. Significant clusters in pre-defined ROIs for the Warning > Safe contrast during the training phase. Clusters are presented at $p_{uncorr.} < .001$. Yellow denotes higher activations for healthy participants, cyan denotes higher activations for the patient groups. Magenta denotes higher activation for opioid-addicted patients compared to non-addicted patients.

For the devaluation test, comparing healthy participants to patients using the Valued > Safe contrast, ROI analyses revealed no clusters significant at FWE < .05. At $p_{uncorr} < .001$, healthy participants showed a significantly higher activation in the left preSMA (xyz = -4 20 62, $T = 3.3$, $p_{uncorr} = .001$) and in the left

anterior putamen (xyz = -22 12 -10, T = 3.2, p_{uncorr} = .001). When comparing opioid-addicted to non-addicted patients, ROI analyses also revealed no clusters significant at either FWE < .05 or at $p_{uncorr.}$ < .001.

Task 2 - Aversive contingency degradation

Groups differed in how unpleasant they rated the aversive stimulus ($F_{2,69}$ = 3.8, p = .028, η_p^2 = .10), with post-hoc tests showing that opioid-addicted patients rated the unpleasantness of the crying babies higher than healthy participants (t_{69} = -2.7, p_{corr} = .024, d = -0.80). Repeated measures ANOVA revealed no group differences in response rates to the different conditions as all three groups reduced their response rates following changes in outcome contingencies (main effect of condition: $F_{2,138}$ = 48.1, $p_{corr.}$ < .001, η_p^2 = .41; group-by-condition interaction: $F_{4,138}$ = 1.2, $p_{corr.}$ = .297; see Figure 19a). Planned contrasts confirmed no group effects for the difference between non-degraded and fully degraded blocks when comparing healthy participants to patients (t_{69} = 0.7, p = .473) or when comparing opioid-addicted to non-addicted patients (t_{69} = -1.5, p = .141). Including the unpleasantness rating as a covariate did not affect the results. Analysis of causality ratings following the blocks showed similarly lower ratings for the partial and full degradation compared to the non-degraded condition across all three groups (main effect of condition: $F_{2,138}$ = 13.6, $p_{corr.}$ < .001, η_p^2 = .16; group-by-condition interaction: $F_{4,138}$ = 1.4, $p_{corr.}$ = .244; see Figure 19b).

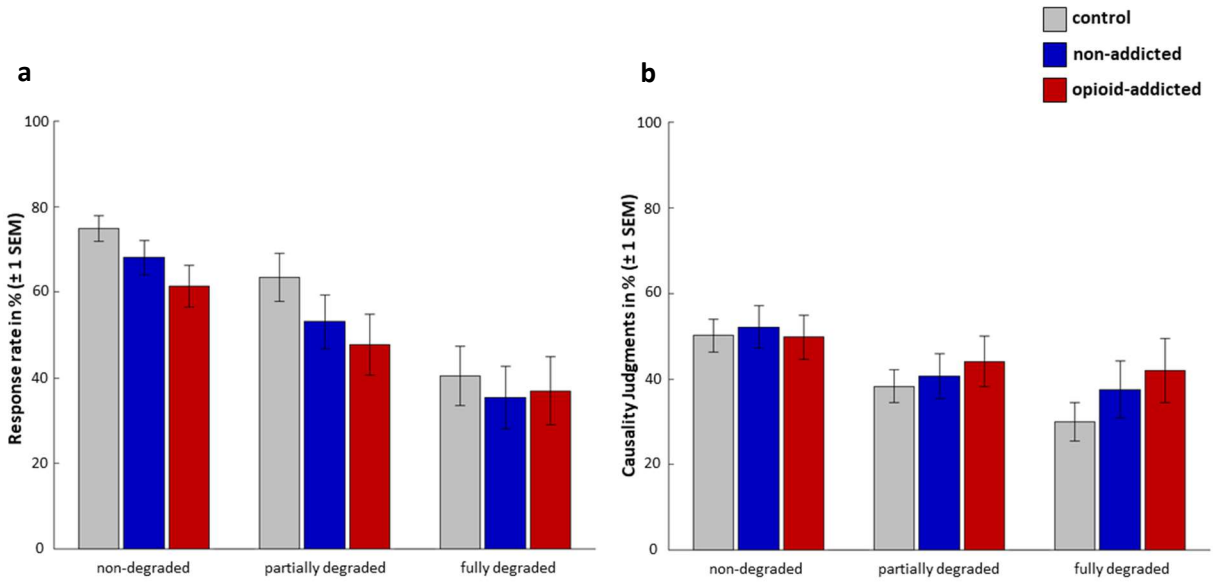


Figure 19. Response rates (a) and causality judgement ratings (b) for the aversive contingency degradation task

Task 3 - Appetitive contingency degradation

Groups did not differ in their self-reported motivation to pick up a 50ct coin from the street ($F_{2,83}$ = 2.2, p = .120). Similar to the aversive task version, there was no main effect of group ($F_{2,83}$ = 1.1, p = .334) and no group-by-condition interaction ($F_{2,164}$ = 0.6, $p_{corr.}$ = 0.649) as all groups adapted their response rates following contingency degradation, indicated by a main effect of condition ($F_{2,163}$ = 38.7, p_{corr} < .001, η_p^2 = .32; see Figure 20a). Response rates were lower following partial degradation compared to the non-degraded blocks (t_{82} = 3.5, p < .001, d = 0.34) and lower following full degradation compared to partial degradation (t_{82} = 6.6, p < .001, d = 0.48). Planned contrasts on the group-by-condition interaction demonstrated similar differences between non-degraded and fully degraded blocks when comparing healthy participants to patients (t_{82} = 0.7, p = .494) or when comparing the two patient groups (t_{82} = 0.4, p = .707). Results did not change when including motivation for monetary rewards as a covariate. A significant interaction effect between group and condition indicated differences in self-reported causality judgments between the groups ($F_{4,163}$ = 3.5, p_{corr} = .009, η_p^2 = .08). Post-hoc tests revealed that healthy participants rated causality on fully degraded blocks significantly lower compared to non-addicted patients (t_{82} = -1.8, p = .043, d = -0.79; see Figure 20b).

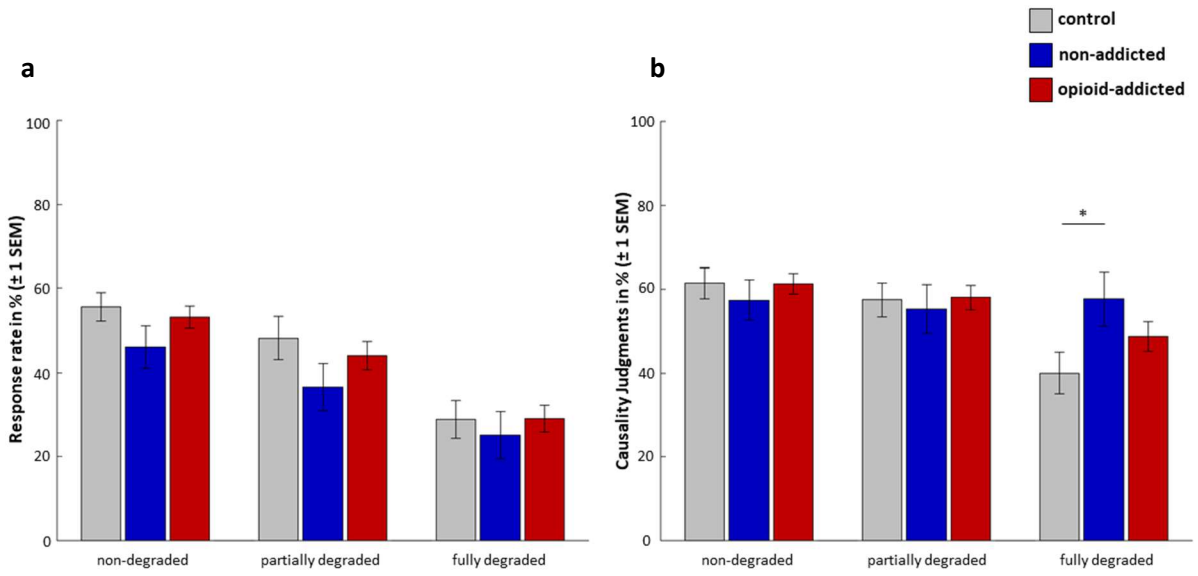


Figure 20. Response rates (a) and causality judgement ratings (b) for the appetitive contingency degradation task

Associations with opioid use characteristics

Task 1: For the non-addicted patients, age of first opioid use was negatively associated with the individually calibrated shock intensity (r = -.60, p = .019). For the opioid-addicted patients, age of first opioid use was differentially correlated with activation in the aforementioned caudate cluster during

training, with younger age of onset related to lower discrimination between warning and safe stimuli in the first block (r = .77, p = .043) and higher discrimination in the second block (r = -.82, p = .250).

Task 2: For the non-addicted patients, daily opioid dose was negatively associated with the avoidance response rate in the non-degraded blocks (r = -.44, p = .039), while age of first opioid use was positively associated with response rate in the fully-degraded condition (r = .43, p = .043). No behavioural measures were significantly correlated with opioid-use characteristics in opioid-addicted patients.

Task 3: For the non-addicted patients, duration of opioid use was negatively correlated with reward response rates in the partially and fully degraded blocks (both r = -.43, p = .034). Again, no significant correlations emerged for opioid-addicted patients.

8.4 - Discussion

In this chapter, I characterised habitual avoidance following chronic opioid use. For this, I employed two different experimental paradigms, in which changes in either outcome value (task 1) or action-outcome relationship (task 2) introduced conflicts between goal-directed and habitual avoidance.

Contrary to my first hypothesis, I did not find convincing evidence for an increased tendency towards avoidance habits in opioid-addicted compared to non-addicted patients. In the first of the two tasks, participants learned to avoid electrical shocks delivered to either wrist until one of the shock electrodes was disconnected. Opioid-addicted patients did not significantly differ from non-addicted patients in that they reduced their response rate towards the stimulus associated with avoiding the shock from the disconnected electrode, indicating of goal-directed behaviour. However, when compared to healthy participants, patients irrespective of addiction status showed a trend towards a lower reduction in response rates following outcome devaluation, which might suggest an increased reliance on outcome-insensitive habits. This is underscored by lower neural activation of the anterior putamen compared to healthy participants during the devaluation test, a brain area associated with goal-directed behaviour (De Wit, Watson, et al., 2012). Furthermore, during avoidance acquisition patients showed decreased avoidance accuracy compared to healthy participants, accompanied by lower activations of the anterior putamen and vmPFC, the connectivity between which is thought to underlie action selection based on outcome values and significantly predicts successful goal-directed behaviour (De Wit, Watson, et al., 2012). Instead, patients showed hyperactivation of the posterior putamen during the training phase, the brain region most widely associated with habits (Tricomi et al., 2009). While these results suggest a bias towards avoidance habits in patients following chronic opioid use irrespective of opioid addiction, it is worth considering that the difference in response rates between stimuli associated with valued and devalued stimuli was almost three times higher in non-

addicted compared to the opioid-addicted patients, indicating a higher reliance on avoidance habits in opioid addiction. Statistical non-significance is potentially related to small patient sample sizes in this task and high within-group variance. To further complicate matters, early instrumental avoidance learning (marked by a significant improvement in avoidance accuracy between the first and second block of training) was apparent in opioid-addicted but not in non-addicted patients. Together with higher activation of the caudate during training, a brain area critical for goal-directed action-outcome learning (De Wit, Watson, et al., 2012; Tanaka et al., 2008), this implies intact goal-directed learning in opioid-addicted patients. Thus, it is conceivable that both patient groups show different mechanistic alterations both resulting in somewhat reduced adjustments to outcome devaluation. In non-addicted patients, impaired goal-directed learning could result in weaker action-outcome associations, while either increased habit formation or a deficit in the ability to exert goal-directed control over habits could underlie lower adjustments in opioid-addicted patients. However, it is important to consider that the finding of a lower differences in response rates following outcome devaluation in patients compared to healthy controls only shows a trend towards statistical significance and given the small sample sizes, it is not possible to draw final conclusions about habitual avoidance in patients with and without opioid addiction from this data. Sufficiently powered replication is necessary to paint a clearer picture of group differences in the interplay between goal-directed and habitual avoidance.

In the second task, participants learned to avoid an aversive auditory stimulus (the sound of crying babies) by pressing a button. Following avoidance training, the causal relationship between behaviour and outcome was decoupled, so that absence of button presses could also lead to avoidance of the outcome. All three groups showed similar reductions in their causality judgement ratings, indicating lower beliefs that the button press causally resulted in avoidance. Importantly, all three groups reduced their response rate in accordance with the lowered causality rating, indicating intact goal-directed behaviour in all groups, with no indication of increased reliance on avoidance habits in opioid-addicted patients. In contrast to both the avoidance training in the Task 1 and the results from Chapter 6, patients showed similar instrumental avoidance compared to healthy participants, indicated by similar response rates during non-degraded blocks. This is likely caused by the task design: Only a single action-outcome association had to be learned, which could favour patients as opioid-addicted patients had been previously demonstrated to only show impaired avoidance learning when a higher number of associations had to be maintained simultaneously (Myers et al., 2017). It would be possible that working memory detriments, demonstrated in chronic opioid users (Baldacchino et al., 2012, 2019), impact more complex learning tasks, but do not affect patients in simple action-outcome learning (Collins & Frank, 2012). It is also possible that the social nature of the aversive stimulus resulted in increased avoidance motivation in patients, especially in opioid-addicted patients, who rated the stimulus as highly unpleasant, positively correlated with initial avoidance responses. Such a valence-

behaviour relationship was notably absent in other tasks reported in this thesis using non-social aversive stimuli.

My secondary hypothesis was that the proposed imbalance between goal-directed and habitual behaviour in opioid addiction would be specific to the aversive domain. I therefore tested whether patients would show increased appetitive habits in a contingency degradation task based on reward-learning which mirrored the aversive version discussed above. Both opioid-addicted and non-addicted patients reduced their response rate following contingency degradation to a similar degree as healthy participants, indicating goal-directed approach behaviour. Interestingly, non-addicted patients reduced their response rate in the fully degraded condition even though they still believed the button press caused the reward. This dissociation between behaviour and explicit causality judgment, indicative of habits, was exclusive to non-addicted patients using opioids for the treatment of chronic pain and occurred only in the appetitive, but not in the aversive condition, possibly indicating impaired belief updating, which could be related to the impaired cognitive flexibility associated with the chronification of pain (Attal et al., 2014).

To summarise, in this chapter I did not find definitive evidence to support that opioid addiction is associated with a shift towards avoidance habits. While addicted patients showed nominally higher habitual responding compared to non-addicted patients following outcome devaluation, this failed to reach statistical significance. Whether this was caused by a true absence of a group difference or rather caused by low statistical power is unclear. They also displayed intact goal-directed behaviour in a simple active avoidance task in which action-outcome contingencies were degraded. A strong theoretical foundation for a role of avoidance habits in the development of opioid addiction warrants a replication of the present results in a sufficiently powered study sample. As predicted, the balance between goal-directed approach behaviour and appetitive habits following chronic opioid use appears to be intact.

Chapter 9 - General Discussion

Opioids are highly potent pain relievers that exert their effects by changing the perceived aversiveness of subjective states. Chronic use potentially alters how aversion is processed, affecting physical and psychological well-being. In addition, following repeated use some develop addiction, a disorder characterised by continued use despite considerable adverse consequences. Inter-individual differences in aversion processing could either increase the risk of developing addiction or serve as protective factors against it. In this thesis, my aim was to characterise aversion processing following chronic opioid use and specify how addicted patients differ from non-addicted patients. Over the course of the previous chapters, I investigated various aspects of aversion processing in chronic opioid users, analysing behaviour on seven different experimental tasks as well as psycho-physiological and neuroimaging data. In this concluding chapter, I will first briefly highlight key results of the study and then integrate the findings in a way that allows to test my overarching hypotheses. Following this, I will discuss strengths and limitations of the present study and provide some thoughts on possible clinical implications of the findings. I will close with a short outlook on future research and some concluding remarks.

9.1 - Summary of key findings

In Chapter 4, I tested whether recognising aversive emotions such as anger and fear differs between patients with chronic opioid use and healthy participants. In contrast to my hypothesis, I did not find group differences in overall or emotion-specific recognition ability, indicating that chronic opioid use does not necessarily lead to impairments in processing social stimuli, aversive or otherwise. Instead, there was an unexpected difference in recognising fearful expressions between the two patient groups, in that opioid-addicted patients required a higher emotional intensity for fear recognition.

In Chapter 5, participants underwent functional MRI while completing a paradigm designed to investigate Pavlovian fear and safety learning. I found no physiological stimulus discrimination in patients, suggesting impaired fear conditioning following chronic opioid use. Importantly, responses to the unconditioned stimulus, an individually calibrated unpleasant electric shock, did not differ between the groups. Neuroimaging demonstrated paradoxically overlapping activation of brain networks implicated in fear and safety processing in patients.

Chapter 6 applied a novel two-part aversive learning paradigm testing instrumental avoidance learning while simultaneously analysing the influence of prior fear conditioning on avoidance behaviour. Results from fear conditioning provided a conceptual replication of the findings in Chapter 5, as both patient groups failed to show psycho-physiological stimulus discrimination despite intact responses to the unconditioned stimulus (in this case a combination of aversive pictures and sounds). As

hypothesised, patients showed impaired instrumental avoidance learning, indicated by a lower avoidance accuracy compared to healthy participants. Moreover, opioid-addicted patients only successfully learned from stimuli that had been previously paired with aversive outcomes, while non-addicted patients showed an opposing pattern.

In Chapter 7, I compared the magnitudes of drug-related and non-drug-related Pavlovian-to-Instrumental-Transfer and found a higher difference between the two in opioid-addicted patients compared to non-addicted patients. More specifically, presenting images depicting opioid use that had previously been paired with monetary gain biased addicted patients' behaviour towards Go responses on trials on which abstaining from responding would have been the optimal choice.

In Chapter 8, participants completed three different experimental tasks to test a proposed shift towards habitual behaviour proposed to underlie the development of addiction. Two separate tasks assessed the balance between goal-directed and habitual avoidance behaviour, and while responses to an outcome devaluation procedure did show nominally higher habitual responding in opioid-addicted patients, the effect was not significant. Importantly, behavioural and fMRI results relied on a very small subgroup of patients, so that the results remain inconclusive. As a side note, in accordance with the results from Chapter 6, instrumental avoidance learning was significantly attenuated in both patient groups compared to healthy participants. A second task applied a contingency degradation paradigm to test for habitual avoidance behaviour and found no evidence for group differences. Interestingly, opioid-addicted patients rated the unpleasantness of a socially relevant aversive stimulus utilised in this task (the sound of crying babies) as significantly higher than non-addicted or healthy participants. A third task, included to test valence-specific habit formation, focused on the balance between appetitive goal-directed and habitual behaviour and did also not yield any significant differences between the three groups.

9.2 - Hypothesis 1: Altered aversion processing following chronic opioid use

The findings of this thesis are in partial agreement with my first hypothesis as some, but not other, aspects of aversion processing were altered in patients with chronic opioid use. Notably, subjective and physiological responses towards aversive stimuli were largely comparable between groups. Participants reported similar levels of unpleasantness when rating aversive pictures. Skin conductance responses, an implicit indicator of sympathetic arousal (Critchley, 2002), demonstrated similar psycho-physiological reactivity towards aversive pictures and sounds across groups. These results add to a heterogeneous literature regarding the responsiveness towards aversive stimuli following chronic opioid use. Patients with opioid addiction have been shown to have higher (Aguilar De Arcos et al., 2008; Gerra et al., 2003), equally high (Biernacki et al., 2018; Walter, Wiesbeck, et al., 2011), and lower

(Smoski et al., 2011; Walter, Degen, et al., 2011; Z.-X. Wang et al., 2010) self-reported affective, physiological or neural responses towards aversive stimuli compared to healthy participants, suggesting that a variety of factors such as type of stimulus, current medication, and co-morbidities contribute to alterations in aversion processing. In this thesis, patients and healthy participants also showed comparable physiological and neural responses towards electric shocks. Groups also perceived similar intensities of electric shock as "aversive but not painful", which was unexpected as previous research indicated lower pain thresholds in opioid-addicted patients as well as in patients with chronic opioid use (OIH; Higgins et al., 2019; Trøstheim & Eikemo, 2024). However, while OIH has been robustly demonstrated to produce lower pain thresholds in the cold pressor test (participants hold their hand in 1°C water and report how soon they experience pain), electrical stimulation has been found to be unaffected by OIH in opioid-addicted and non-addicted patients (Doverty et al., 2001; Hay et al., 2009), potentially due to different underlying neurophysiological pathways (thermal nociceptors and C-fibres for cold pressor pain, Aδ-fibres for electrical stimulation). The present data supports this distinction.

Interestingly, while patients as a whole did not differ from healthy participants in recognising aversive facial expressions such as angry or fearful faces, opioid-addicted patients required a higher intensity of emotional expression for fear recognition compared to non-addicted patients. This dovetails previous findings of diminished fear recognition in addicted patients (Bland & Ersche, 2020; Y.-T. Kim et al., 2011; Trick et al., 2014). While this specific impairment could be a consequence of addiction-related neuroadaptations in limbic brain structures underlying the processing of fearful expressions (Adolphs et al., 1994; Zhao et al., 2013) or caused by drug-use related alterations in endocrine function (Ersche et al., 2015), it is also possible that it predates the onset of opioid use, in which case it might increase the risk of developing addiction. Difficulties in recognising fearful faces has been related to alexithymia (Craparo et al., 2016), impaired socio-cognitive skills (Corden et al., 2006) and antisocial personality traits (Marsh & Blair, 2008), all of which have been linked to opioid addiction (Compton et al., 2005; Honkalampi et al., 2022; McDonald et al., 2013). Additionally, opioid-addicted patients, but not non-addicted patients, rated the sound of crying babies as significantly more aversive compared to healthy participants, with many addicted patients rating close to the scale maximum ("extremely unpleasant"). This finding suggests a heightened reactivity towards aversive social stimuli, which, together with reports of stronger feelings of social rejection in opioid-addicted patients (Bach et al., 2019; Kroll et al., 2019) aligns with the notion that high sensitivity towards social stress is a key driver underlying addiction (Heilig et al., 2016).

Aversive learning was found to be disrupted in patients with chronic opioid use. In two separate fear conditioning tasks, patients displayed a lack of stimulus discrimination indicative of impaired Pavlovian learning. This agrees with a previous study demonstrating diminished fear conditioning in opioid-

addicted patients (Basden et al., 2016) and extends these findings to non-addicted patients. Fear conditioning depends on processing aversive prediction errors (S. S. Y. Li & McNally, 2014; Spoormaker et al., 2011), which are likely encoded in μ-receptor-rich brain regions such as the insula and the PAG. Structural changes and receptor desensitisation in these areas have been observed following chronic opioid use (Bach et al., 2021; Bagley et al., 2005; Wollman et al., 2017), potentially disrupting efficient functioning of the neural circuits underlying fear conditioning. The fMRI results reported here demonstrate that some brain areas typically implicated in fear learning (Fullana et al., 2016) showed inadequately high activation in response to safety signals (such as the caudate), suggesting disorganised neural responses to threat and safety. Interestingly, other fear-related regions such as anterior and posterior insula correctly tracked threat signals in patients. While the posterior insula processes the intensity of aversive stimuli (Segerdahl et al., 2015), the anterior part tracks aversive prediction errors (Hoskin & Talmi, 2023) and pain-related expectations (Fazeli & Büchel, 2018), which suggests that the differential pairing of conditioned and unconditioned stimuli was successfully registered. This leads to the conclusion that components of fear learning other than the processing of aversive prediction errors might be impaired following chronic opioid use. For example, associability, i.e. the extent to which a stimulus is expected to contribute to future learning (J. Li et al., 2011), gates attention towards certain stimuli and determines whether prediction errors are used to update outcome values. Associability is encoded in the amygdala (J. Li et al., 2011), a brain region reduced in volume following repeated opioid use (Lin et al., 2016; Younger et al., 2011). In the present study, patients showed reduced left amygdala activation during fear conditioning. If amygdala structure and functioning were disrupted following chronic opioid use, insufficient attribution of associability to the safety signal could explain why patients failed to reduce fearful responses, which would then result in defunct stimulus discrimination. Ultimately, more in-depth computational analyses are warranted to fully understand which mechanisms underlie the impaired fear conditioning following chronic opioid use. Interestingly, in the non-addicted patient group, longer duration of opioid use was associated with smaller physiological stimulus discrimination in both fear conditioning tasks, suggesting learning deficits as a consequence of altered opioidergic function due to chronic use. The absence of such a relationship in opioid-addicted patients could either imply that learning impairments predated opioid use or that they are caused by addiction-related neuroadaptations. However, since the size of the subsample on which this analysis was based was rather small, it is also possible that duration of use is indeed associated with fear conditioning deficits in opioid-addicted patients, but statistical power was insufficient to detect it.

In addition to the observed impairments in fear conditioning, patients with chronic opioid use showed attenuated instrumental avoidance learning. This extends previous findings of reduced learning from negative outcomes in opioid-addicted patients (Gradin et al., 2014; Myers et al., 2017) and

demonstrates that this result is not limited to opioid addiction but also present in non-addicted patients. Further solidifying this relationship, impaired avoidance learning in patients was observed in two separate paradigms despite pronounced differences in task design: While the first was a probabilistic learning task in which participants avoided monetary loss, the second resembled classic active avoidance paradigms (Krypotos, 2015) in which participants learned to avoid electric shocks (as training in a task testing avoidance habits). Comparison of avoidance accuracy of healthy participants in the two tasks suggested higher task difficulty in the former (mean accuracy: ~ 67%) compared to the latter (mean accuracy: ~ 90%), which might be relevant as opioid-addicted patients differed from healthy participants only at higher task difficulties in a previous study (Myers et al., 2017). In this thesis, results were similar between tasks: In both tasks, patients did learn to avoid aversive outcomes but did so with lower accuracy compared to healthy participants. While this underscores the robustness of the finding, other results from this thesis demonstrate its specificity. In an avoidance task that required very little learning, avoidance response rates were similar between the three groups, showing that not avoidance behaviour per se is reduced but rather avoidance learning. Additionally, while not the primary focus of this task, behaviour in the instrumental learning phase of the PIT paradigm revealed no group differences in accuracy for instrumental reward learning, further demonstrating that deficits are specific to aversive learning instead of instrumental learning in general. One possible explanation for this key result is that avoidance learning hinges on successful fear conditioning (Mowrer, 1951). Following this logic, avoidance accuracy would be attenuated due to the defunct fear conditioning in the two patient groups, as people only avoid what they have learned to fear. Opposing this notion, however, I found no significant correlations between stimulus discrimination during fear conditioning and avoidance accuracy during avoidance learning for any group. While the two learning mechanisms are without doubt intertwined and share an overlapping neurobiology (Fullana et al., 2016; H. Kim et al., 2006), more recent theories postulate a complex relationship between the two, with cognitive constructs such as outcome expectancy crucially modulating avoidance behaviour (Krypotos, 2015; Lovibond et al., 2008). This is related to contemporary accounts of both aversive learning mechanisms that stress the importance of learning from violated predictions (Maia, 2010). In instrumental learning, predicted outcome values of competing behavioural options are weighed against each other and the option with the lowest aversive value is selected. Fear conditioning fMRI findings discussed in the previous paragraph suggest that the tracking of aversive values is intact in patients with chronic opioid use and that other processes instead cause attenuated aversive learning. Neuroimaging results from a study investigating avoidance learning in opioid-addicted patients showed reduced ventral striatum activation when encoding avoidance (i.e. positive) prediction errors (Gradin et al., 2014). This could mean that learning from negative outcomes is indeed intact, but learning from avoiding these outcomes is not. Additionally, during decision-making patients showed

reduced activation of the inferior frontal gyrus (Gradin et al., 2014), associated, among others, with switching from initial exploratory to exploitative choice behaviour (Domenech & Koechlin, 2015). Diminished function in this brain area could imply that, despite correct value computation, action selection remains explorative in patients, favouring more random choices and explaining the sub-optimal choice behaviour. Latent parameters such as choice randomness or different learning rates from positive and negative prediction errors could be quantified by fitting generative models on the behavioural data, which might help to clarify the mechanisms underlying impaired avoidance learning.

Intriguingly, further analysis revealed that patient groups differed in how previous fear conditioning impacted instrumental learning, in that opioid-addicted patients only learned to avoid money loss from stimuli that had previously been paired with aversive material. One possible explanation for this could be that fear conditioning increased the aversive value of these stimuli, promoting avoidance of these stimuli (conditioned avoidance; Maia, 2010). If that was the case, however, one would expect higher avoidance rates for the aversively conditioned stimuli in early trials, indicative of a general avoidance bias, instead of the finding that addicted patients significantly improved their accuracy only in regard to these stimuli. Rather, this could be explained by the increased salience of the aversively paired stimuli enabling learning. In probabilistic instrumental learning environments, efficient learning from prediction errors is only possible when stimuli possess sufficient salience (Boehme et al., 2015) and it is possible that opioid-addicted patients require highly salient stimuli for effective value updating, potentially due to addiction-related aberrant salience attribution when non-drug-related stimuli are concerned (Zilverstand et al., 2018). Non-addicted patients, on the other hand, displayed an opposing pattern, only learning from stimuli that were not previously used in fear conditioning. More in-depth analysis is warranted to better understand which mechanistic processing underlies the differential learning profiles in the two patient groups.

To summarise, while the processing of non-social aversive stimuli remained unchanged, aversive learning was attenuated following chronic opioid use. In addition, the processing of aversive social stimuli differed between patients with and without opioid addiction.

9.3 - Hypothesis 2: Specific impairments in regulatory control over avoidance behaviour

The findings of this thesis are in partial agreement with my second hypothesis. In a Pavlovian-To-Instrumental Transfer paradigm, the presentation of drug-related stimuli previously paired with reward biased decision-making towards sub-optimal approach behaviour in addicted, but not in non-addicted, patients. The results are in line with the incentive sensitisation theory (Robinson & Berridge, 1993), which states that the strong hedonic impact ("liking") of addictive drugs results in high incentive salience ("wanting") of drug-use related stimuli. Incentive salience is an adaptive mechanism directing

attention and motivation towards stimuli associated with pleasant states (Olney et al., 2018); however, fuelled by the pharmacological properties of psychoactive drugs (S. Peciña & Berridge, 2013), drug-related stimuli can induce a strong approach bias, contributing to the development of compulsive drug use. PIT presents an opportunity to experimentally assess incentive salience (Olney et al., 2018), which to date has not been tested in opioid-addicted patients. The findings of increased drug-related PIT presented in this thesis therefore provide novel evidence for incentive sensitisation in opioid addiction. The absence of higher drug-related compared to non-drug-related PIT in non-addicted patients suggests this effect as an important distinction between the groups. If encountering opioid use-related stimuli elicits strong "wanting" in some opioid users, but not in others, this would increase the risk of the former and decrease the risk of the latter group to use opioids despite knowledge of potential detrimental consequences. Pronounced drug "wanting" is often equated with craving[8], a preoccupation with drug use that often results in relapse (Vafaie & Kober, 2022). The relationship between craving and relapse has been repeatedly demonstrated in opioid-addiction (McHugh et al., 2014, 2016; Murphy et al., 2018), and while factors other than exposure to opioid use-related stimuli can also increase craving (such as stress; Hyman et al., 2007; Preston et al., 2018), drug-related PIT could represent an underlying neurocognitive mechanism. Notably, the magnitude of behavioural and neural PIT predicted relapse in detoxified patients addicted to alcohol (Garbusow et al., 2016; Sekutowicz et al., 2019) and it is possible that a similar relationship could be found in opioid addiction. The finding that craving is an important predictor for future opioid use in chronic pain patients (Wasan et al., 2009) further underscores the idea that high opioid "wanting" could help to identify who is at-risk to develop difficulties maintaining control over opioid use. Before, however, the reliability of the findings presented in this thesis have to be tested and longitudinal measures have to clarify the potential relationship between PIT and prospective drug-use related measures.

Contrary to my hypothesis, I did not find any conclusive evidence for an imbalance between goal-directed and habitual control over avoidance behaviour in opioid-addicted patients. The habit theory of addiction (Everitt & Robbins, 2005, 2016) proposes that a gradual shift in regulatory control promotes the development of compulsive drug use, with use progressing from goal-directed behaviour into a stimulus-response habit. While evidence suggests this to be the case for cocaine addiction (Ersche et al., 2016, 2021), this has not yet been tested in opioid-addicted patients, where, contrary to the rewarding effects of psychostimulants, use is thought to be driven by negative reinforcement (Koob, 2020; Pantazis et al., 2021). This suggests an over-reliance on habitual avoidance as a driving factor behind the development of opioid addiction (see also Ersche et al., 2016; Robbins, 2019). Here, I tested the balance between goal-directed and habitual avoidance in different tasks designed to

[8] Although some authors note that „wanting" represents a rather short-lived desire that can easily be ignored, while "craving" is more intense and intrusive (e.g. DiFranza, 2016).

induce conflict between the two. In the first task, participants learned to avoid electric shocks to their left or right wrist before one of the two shock electrodes was disconnected. Following this procedure, opioid-addicted patients reduced their response behaviour only by 11% compared to a 31% reduction in non-addicted patients and a 44% reduction in healthy participants. While this difference is indicative of habitual avoidance, it failed to reach statistical difference. Crucially, this analysis only included data from seven addicted and fifteen non-addicted patients, severely limiting its interpretability due to a lack of statistical power. In a similar vein, neuroimaging findings from this task have to be taken in with caution. While I found no differences in neural activation between the two patient groups during the habit test, opioid-addicted patients reported higher activation in the caudate (associated with goal-directed behaviour) and lower activation in the preSMA (associated with habits; De Wit et al., 2012) during avoidance training, contrasting behavioural results. Ultimately, the small sample size of the dataset does not allow for any strong claims regarding neural substrates of goal-directed and habitual avoidance, warranting replication in a sufficiently powered sample. A second behavioural task could draw on sufficient sample sizes and utilised a contingency degradation paradigm, in which performance of avoidance behaviour is gradually decoupled from its outcome. In this novel task, participants across groups correctly tracked the diminishing causal relationship between action and outcome and adapted their avoidance behaviour accordingly, indicating goal-directed behaviour. Multiple explanations for the unexpected absence of increased habitual tendencies in opioid-addicted patients are conceivable. Predominance of habitual avoidance could be state-dependent, meaning that it would only be apparent in certain (albeit critical) situations. To promote compulsive opioid use, it would be sufficient if regulatory control over avoidance behaviour shifted towards habits in states of withdrawal. As long as patients' opioidergic tone is high, their behaviour could be goal-directed, but once the low opioidergic tone and high stress levels – both factors that facilitate habitual behaviour (Schwabe et al., 2012; Voon et al., 2020; Wassum et al., 2009) - associated with opioid withdrawal set in, behaviour could be governed by stimulus-response habits, prompting drug use. This could explain why, in the present study, patients on OMT with no signs of withdrawal did not show habitual avoidance. Incidentally, entering the MRI environment can induce physiological and subjective stress (Tessner et al., 2006), which could be why I found a trend towards group differences in the first task but not in the behavioural task conducted outside of the MRI environment. It is also possible that the absence of the hypothesised findings is due to the subgroup of opioid-addicted patients included in the present study. The majority (87%) reported no present use of illicit opioids, with an average self-reported abstinence from heroin of over 10 years. While opioid addiction is a chronic disorder and relapse after prolonged abstinence is not uncommon (Strang et al., 2020), it is possible that patients included in this study have re-gained control over opioid use, an assumption supported by relative low scores on a questionnaire assessing compulsive drug use compared to other studies in OMT patients

(e.g. Blanken et al., 2012; Chang & Peters, 2023). Hence, it is possible that opioid-addiction is indeed associated with increased habitual tendencies, but patients can learn to overcome these tendencies, allowing them to abstain from use. Lastly, it is important to acknowledge the possibility that opioid addiction is, in fact, not characterised by increased habitual avoidance. An overreliance on habits has been previously demonstrated in addiction to cocaine (Ersche et al., 2016, 2021) but not, for example, nicotine (Luijten et al., 2020), with mixed findings for alcohol (Sjoerds et al., 2013; Van Timmeren et al., 2020). Psychoactive drugs vastly differ in their pharmacological profiles, with particularly opposing effects between opioids and psychostimulants such as cocaine (Badiani et al., 2011). Cocaine use rapidly increases extra-cellular dopamine (Volkow et al., 1997), the neurotransmitter primarily implicated in habit formation (Belin & Everitt, 2008; De Wit, Standing, et al., 2012), thus potentially facilitating habitual behaviour. While dopaminergic neurotransmission likely also contributes to the effects of opioids, its role in opioid addiction is less clear-cut compared to stimulants (Badiani et al., 2011; Daglish et al., 2008). Lower dopaminergic activity following opioid use compared to stimulant use could therefore potentially explain the difference in habit formation.

To summarise, while regulatory control was not found to favour habitual over goal-directed avoidance in opioid-addicted patients, presentation of drug-related Pavlovian stimuli paired with monetary gain induced an approach bias into patients' behaviour.

9.4 - Strengths and limitations

This study utilised a comprehensive set of behavioural tasks that allowed to systematically characterise aversion processing following chronic opioid use. Some of the tasks applied in this thesis have been previously validated in patient samples, while others are unpublished, contributing to the novelty of the findings. Notably, the behavioural constructs under investigation were partly overlapping between tasks, enabling conceptual replication of the underlying mechanisms, (i.e. testing of the same psychological phenomenon in a different parameter space; Derksen & Morawski, 2022). For example, the finding of impaired fear conditioning in patients was initially derived from a task in which pictures of angry faces were partially reinforced with aversive electric shocks. The fact that the same result was also obtained in a different task using deterministic pairing of coloured patterns with aversive images and sounds underscores its robustness. In similar fashion, the finding of impaired avoidance learning could be conceptually replicated in a separate task using different stimuli and behavioural outcome measures. Moreover, while the focus of the study concerned aversion processing, some tasks also tested aspects of reward processing (such as instrumental reward learning or appetitive habit formation), which allowed to contrast aversion- and reward-related results. This increased the specificity of the findings, such as identifying impaired instrumental avoidance but intact instrumental

reward learning in patients. Functional neuroimaging and skin conductance responses provided additional data sources, offering context when interpreting behavioural results. Another strength of the study consisted of the three-group design, which included opioid-addicted patients, non-addicted patients, and healthy participants. Most studies investigating chronic opioid use in the past compared one patient group to a healthy control group, often leading to ambiguous results where it is unclear whether chronic opioid use is directly related to the observed effect, whether it might have predated use, or whether third variables explain the findings. While this study is not the first to compare opioid-addicted patients to non-addicted patients (e.g. Baldacchino et al., 2015, 2019), it is the first to do so while decidedly investigating aversion processing, a construct central to understanding opioid use trajectories. Additionally, patients included in this study were carefully selected to ensure that opioid-addicted patients had no current stimulant use and had at no point in their life been addicted to stimulants. Studies of avoidance behaviour in patients with stimulant addiction and co-morbid opioid addiction revealed patterns similar to patients only addicted to stimulants (Ersche et al., 2016; Lim et al., 2019), suggesting that high stimulant use might overshadow more subtle effects pertaining to opioid addiction. Therefore, the present sample provides a rare opportunity to observe psychological and neural correlates of chronic opioid use isolated from psychostimulant use.

The aforementioned inclusion of non-addicted patients with chronic opioid use is a major strength of this study. However, it has to be acknowledged that patients from this group are currently in treatment for a number of severe pain syndromes. Chronic pain is associated with changes in attention, motivation, mood, and learning (Simons et al., 2014) and has also been linked to altered aversion processing, such as fear conditioning (Meulders et al., 2015; Schlitt et al., 2022) and aversive Pavlovian-To-Instrumental Transfer (Nees et al., 2020). Moreover, escalating avoidance behaviour, proposed in this thesis as a crucial variable in the development of opioid addiction, is also central to a family of theories regarding the chronification of pain (so called fear-avoidance models; Vlaeyen et al., 2016). To further complicate matters, different chronic pain syndromes have been linked to different patterns of aversion processing (e.g. Icenhour et al., 2015; Jenewein et al., 2013). This encumbers differentiating which of the alterations observed in this thesis are directly related to opioid use and which pertain to the presence of chronic pain. Correlations of task measures with opioid use characteristics such as duration of use can serve as indicators but can also be confounded by pain characteristics such as duration of chronic pain, which was not assessed in the study. Since no chronic pain patients without opioid use were included in this study, it is impossible to make specific claims regarding opioid use, and the rather small sample size did not allow for subgroup analyses into different chronic pain syndromes. The same is the case for the different types of opioids prescribed to patients, which differ in their pharmacological profiles (e.g. methadone acts as a full μ-agonist, buprenorphine as partial μ-agonist and κ-antagonist, tapentadol as μ-agonist and noradrenaline reuptake inhibitor), and the

diverse list of adjunct medications prescribed in both groups (e.g. selective serotonin reuptake inhibitors, GABA agonists) that could possibly influence task outcomes[9]. On another note, all patients in our study have recently used opioids, so it is ultimately not possible to disentangle acute and chronic opioid effects. While some task measures in this thesis were associated with opioid dose and thus might be considered related to acute effects, this is likely confounded by tolerance and, potentially, pain intensity or interference in non-addicted patients. Some study designs have attempted to parse out acute and chronic effects by testing some patients before and some after taking their prescribed medication (e.g. Carlyle et al., 2020; Gradin et al., 2014). Ultimately, one has to acknowledge that aversion processing following chronic opioid use is shaped by a steady and partially idiosyncratic interplay between acute use, within-system tolerance, addiction-related between-systems adaptations, and predispositions. Given the high heterogeneity in opioid use characteristics, the fact that this thesis identified consistent and replicable group effects speaks to the robustness of the findings.

9.5 - Clinical implications

A first encouraging outcome from this study concerns the potential of OMT to reduce harm in patients with opioid addiction. Patients were enlisted in OMT for an average of 14 years and at the time the study was conducted almost all patients had ceased illicit heroin use, with an average time of abstinence of over 10 years. No patients reported any intravenous drug use, associated with the most severe harmful consequences of opioid addiction, including fatal overdoses (Warner-Smith et al., 2002). Additionally, patients reported low levels of craving, even after completing an experimental task where drug-related pictures were displayed, while low scores on the OCDUS indicated low levels of preoccupation with and interference by drug-related thoughts (Franken et al., 2002). Since inclusion criteria of this study introduced a selection bias into the sample, this cannot be taken as evidence for the efficacy of OMT; it does, however, illustrate that at least some patients in OMT are able to deescalate the spiral of opioid addiction.

One key finding from this thesis indicates that, in opioid-addicted patients only, pictures related to opioid use that have been previously paired with pleasant outcomes can increase approach behaviour despite negative consequences. It is well established that exposure to opioid-related stimuli increases craving and risk of relapse (Back et al., 2014; Fatseas et al., 2011) and PIT presents a probable

[9] Serotoninergic transmission especially has been implicated in processing and learning from aversive stimuli (Crockett & Cools, 2015). Additionally, increasing serotonin levels have been associated with reducing compulsive drug use (Y. Li et al., 2021). Chronic opioid use in pain patients has been found to lower activation at serotoninergic receptor sites (Baliki et al., 2024), which warrants further investigation regarding the role of serotonin in the development of opioid addiction.

underlying mechanism (Berridge & Robinson, 2016). While reducing incentive salience towards drug-related stimuli seems like a promising avenue for addiction treatment, one has to proceed carefully. Following cue-exposure therapy, where patients watched videos displaying drug-use on their own, a disproportionally high number of patients with opioid addiction dropped out of treatment and relapsed, compared to an active control group (Marissen et al., 2007). Novel treatment approaches have to make sure to equip patients with methods of coping with states of high drug "wanting". As an example, mindfulness-based treatment approaches aim to prevent relapse by strengthening top-down control over bottom-up urges elicited by drug-related stimuli, redirecting attention away from associated drug effects (Garland, 2024). Mindfulness interventions have been demonstrated to result in decreased physiological and neural responses to opioid-related cues and reduced craving in chronic pain patients with opioid misuse (Garland et al., 2014, 2019). While this is promising, these findings have yet to be extended to patients with opioid addiction.

The finding of impaired avoidance learning in opioid-addicted patients is particularly relevant. It implies that the threat of punitive measures, such as exclusion from a treatment program following relapse, will be largely ineffective given the suboptimal decision-making when it comes to avoiding negative consequences. A similar logic makes it improbable that the risk of draconic legal consequences (associated with "law and order" policies) would successfully deter addicted patients from use. Patients showed no impairments, on the other hand, in instrumental reward learning, indicating that leveraging incentives instead of punishments as motivation could be more effective. Contingency management, where patients receive rewards for meeting certain treatment goals (such as negative urine screens indicating abstinence), has been shown to successfully reduce illicit drug use and improve treatment adherence in opioid-addicted patients in OMT (Bolívar et al., 2021). However, ultimately, it is important to acknowledge that opioid addiction is a complex disorder and incentivising patients against drug use only treats symptoms and not underlying causes.

Non-addicted patients included in this study had been using opioids for an average of 7 years without evidence of misuse, further demonstrating that long-term opioid use alone is not sufficient for developing addiction. Interestingly, while some known risk factors for the development of opioid addiction such as male gender and low formal education (Strang et al., 2020; Webster, 2017) were indeed more prevalent in the opioid-addicted group, other established risk factors such as a family history of drug abuse, severity of childhood abuse or depressive symptoms were equally pronounced in non-addicted patients, indicating a complex interaction between potentially cumulative risk factors. Non-addicted patients reported high intensities of chronic pain despite daily opioid use, underscoring the limited efficacy of opioids in treating chronic non-cancer pain (Ballantyne & Shin, 2008; Reinecke et al., 2015). Together with recent research stating that long-term opioid therapy is associated with

greater pain interference and detriments to emotional functioning in chronic pain patients (Rached et al., 2022), this corroborates that opioids, while effective in treating acute pain, may lead to detrimental outcomes when prescribed long-term. Lastly, the findings of impaired avoidance learning reported here might be relevant in face of the recent popularity of the fear-avoidance model of chronic pain (Vlaeyen et al., 2016). This theory states that patients that tend to avoid pain-related activities enter a self-reinforcing cycle of fear of pain, negative mood, and disability, which favours the chronification of pain, while patients that tend to actively approach their recovery process are able to restore mobility, resulting in higher self-efficacy and positive affect. On this basis, one would perhaps expect disproportionally high avoidance behaviour in patients with chronic pain, which was not the case in this thesis. However, patients displayed reduced avoidance accuracy, potentially indicating sub-optimal decision-making when it comes to avoiding negative outcomes. Experimental paradigms modelling approach-avoidance conflicts (e.g. Pittig et al., 2018) should be well suited to investigate fear-avoidance scenarios related to pain chronification. Additionally, although not statistically significant, non-addicted patients showed a trend towards reduced responsiveness to aversive outcome devaluation, which indicates an increased tendency for avoidance habits. The role of habitual avoidance has not received much attention so far in the domain of chronic pain; following this line of research could be a promising endeavour.

9.6 - Conclusion

This thesis for the first time systematically characterises aversion processing in patients with chronic opioid use. Results indicate deficits in both Pavlovian fear conditioning and aversive instrumental learning in patients with and without opioid addiction, although underlying mechanisms may differ between the two groups. In general, subjective and physiological responses towards aversive stimuli were comparable to healthy participants, but opioid-addicted patients especially showed specific alterations in processing aversive social stimuli, which warrants further study. Moreover, novel and valuable insights emerged into the motivational properties of opioid use-related visual stimuli in addicted patients. The high incentive salience attributed to drug-related pictures has been proposed to promote relapse and thus contribute to the maintenance of addiction. In this thesis, such stimuli elicit a disadvantageous approach bias in addicted, but not non-addicted, patients, which could provide a neurocognitive basis for cue-induced relapse. Future research should address how this relates to prospective drug use outcomes and treatment success. Finally, contrary to expectations, no conclusive evidence for an overreliance on avoidance habits was found in opioid-addicted patients. While methodological constraints could explain this result, it is also possible that chronic opioid use, contrary to regular psychostimulant use, does not facilitate habit formation. In addition, the role of habitual

avoidance in chronic pain warrants future investigation. Taken together, the results of this thesis shed light on the relationship between aversion processing and chronic opioid use and provide multiple avenues for understanding and treating opioid addiction.

Summary

For millennia, humans have used opioids to relieve physical pain, anxiety, and insomnia. Today, opioids often provide a last resort for patients with chronic pain for whom other treatments fail. Repeated use, however, carries a risk of misuse and around 10 to 20 percent of users develop opioid addiction, a chronic disorder characterised by continued opioid use despite severe adverse consequences. Who develops addiction and who does not is a poorly understood, yet critically important question, especially in light of the current opioid overdose epidemic in North America, costing over 100,000 lives per year.

Functionally, opioids exert their effects by stimulating the μ-receptor system which manipulates the perceived aversiveness of states. Chronic use results in neuroadaptations to this system, potentially disturbing aversion processing, an umbrella term for mechanisms related to the experience of disliking a stimulus or state. In this thesis, my aim was to systematically characterise aversion processing following chronic opioid use with and without addiction in order to 1) identify potential alterations related to chronic opioid use and 2) clarify which aspects of aversion processing are specifically altered in addicted patients and might thus contribute to the development or maintenance of opioid addiction. For this, I analysed behavioural, psycho-physiological, and functional neuroimaging data of 86 participants, 23 of whom were addicted to opioids and 25 were not addicted, but were using opioids chronically for the treatment of chronic pain.

While subjective and physiological responses towards aversive stimuli were by and large similar to healthy participants, both patient groups displayed attenuated fear conditioning as well as aversive instrumental learning. Opioidergic neurotransmission underlies aversive prediction error signalling, which suggests that chronic opioid use might lead to the observed learning deficits. However, the cross-sectional nature of this study does not allow for causal inferences on the relationship between chronic opioid use and altered aversive learning. As hypothesised, opioid-addicted patients showed a higher motivational response to drug-related pictures compared to neutral pictures that had previously paired with rewards. Crucially, presentation of drug-related stimuli induced a behavioural bias, promoting approach over avoidance in disadvantageous situations and potentially contributing to relapse in opioid addiction. Contrary to theoretical considerations, there was no conclusive evidence for an overreliance on habits in opioid-addicted patients, indicating intact regulatory control over avoidance behaviour.

In summary, this thesis provides valuable evidence on aversion processing following chronic opioid use and offers novel targets for understanding and treating opioid addiction.

Seit Jahrtausenden gebrauchen Menschen Opioide um Schmerzen, Angst und Insomnie zu lindern. Heutzutage stellen Opioide häufig für Patient:innen mit chronischen Schmerzen eine letzte Chance dar, wenn andere Behandlungsoptionen scheitern. Wiederholter Gebrauch birgt ein gewisses Missbrauchsrisiko und etwa 10 bis 20 Prozent von Menschen, die Opioide gebrauchen, entwickeln Opioidabhängigkeit, eine chronische Erkrankung, die von anhaltendem Gebrauch trotz schwerer negativer Folgen gekennzeichnet ist. Wer abhängig wird und wer nicht ist eine nur schlecht verstandene, aber hochrelevante Frage, gerade vor dem Hintergrund der gegenwärtigen Opioidkrise in Nordamerika, die jedes Jahr über 100 000 Menschenleben fordert.

Opioide erzielen ihre Wirkung, indem sie das μ-Rezeptorsystem stimulieren, das die wahrgenommene *aversiveness* eines Zustands verändert. Chronischer Gebrauch resultiert in Neuroadaptionen und stört möglicherweise *aversion processing*, ein Überbegriff für Mechanismen, die mit dem Nicht-Mögen von Reizen oder Zuständen zusammenhängen. In dieser Thesis war es mein Ziel, aversion processing nach chronischem Opioidgebrauch mit und ohne Abhängigkeit systematisch abzubilden, um 1) potenzielle opioidbezogene Veränderungen zu identifizieren und 2) herauszuarbeiten, welche Aspekte von aversion processing spezifisch bei abhängigen Patient:innen verändert sind und daher zur Entstehung oder Aufrechterhaltung von Abhängigkeit beitragen könnten. Dafür habe ich Verhaltens-, psychophysiologische und funktionale Bildgebungsdaten von 86 Teilnehmer:innen analysiert, von denen 23 opioidabhängig waren und 25 nicht opioidabhängig waren, aber Opioide chronisch zur Schmerzbehandlung einnehmen.

Subjektive und physiologische Reaktionen auf aversive Reize waren im Großen und Ganzen vergleichbar zu gesunden Teilnehmer:innen. Beide Patient:innengruppen zeigten reduziertes *fear conditioning* und aversives *instrumental learning*. Opioiderge Signale bilden die Basis für aversive Vorhersagefehler, was andeutet, dass chronischer Opioidgebrauch zu den beobachteten Lerndefiziten führen könnte; der Querschnittansatz dieser Studie lässt allerdings keine kausalen Schlüsse zu. Wie vorhergesagt zeigten opioidabhängige Patient:innen stärkere motivationale Reaktionen auf opioidbezogene Bilder verglichen mit neutralen Bildern, die zuvor mit einer Belohnung gepaart wurden. Opioidbezogene Bilder lösten einen *bias* aus, der zu Annäherungs- statt Vermeidungsverhalten in unvorteilhaften Situationen führt und möglicherweise Rückfälle begünstigt. Entgegen theoretischer Annahmen gab es keine stichhaltige Evidenz für einen übermäßigen Bezug zu gewohnheitsmäßigen Verhalten bei abhängigen Patient:innen, die intakte Kontrolle über Vermeidungsverhalten zeigten.

Diese Thesis liefert wertvolle Erkenntnisse über aversion processing nach chronischem Opioidgebrauch und identifiziert neue Ansätze, um Opioidabhängigkeit zu erforschen und zu behandeln.

References

Adamson, S. J., Kay-Lambkin, F. J., Baker, A. L., Lewin, T. J., Thornton, L., Kelly, B. J., & Sellman, J. D. (2010). An improved brief measure of cannabis misuse: The Cannabis Use Disorders Identification Test-Revised (CUDIT-R)☆. *Drug and Alcohol Dependence*, *110*(1–2), 137–143. https://doi.org/10.1016/j.drugalcdep.2010.02.017

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68–72. https://doi.org/10.1038/nature03086

Adolphs, R., & Tranel, D. (2003). Amygdala damage impairs emotion recognition from scenes only when they contain facial expressions. *Neuropsychologia*, *41*(10), 1281–1289. https://doi.org/10.1016/S0028-3932(03)00064-2

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669–672. https://doi.org/10.1038/372669a0

Aguilar De Arcos, F., Verdejo-García, A., Ceverino, A., Montañez-Pareja, M., López-Juárez, E., Sánchez-Barrera, M., López-Jiménez, Á., Pérez-García, M., & PEPSA team. (2008). Dysregulation of emotional response in current and abstinent heroin users: Negative heightening and positive blunting. *Psychopharmacology*, *198*(2), 159–166. https://doi.org/10.1007/s00213-008-1110-2

Al-Hasani, R., & Bruchas, M. R. (2011). Molecular Mechanisms of Opioid Receptor-dependent Signaling and Behavior. *Anesthesiology*, *115*(6), 1363–1381. https://doi.org/10.1097/ALN.0b013e318238bba6

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

American Psychological Association. (2018). Aversion. In *APA Dictionary of Psychology*. https://dictionary.apa.org/aversion

Apergis-Schoute, A. M., Gillan, C. M., Fineberg, N. A., Fernandez-Egea, E., Sahakian, B. J., & Robbins, T. W. (2017). Neural basis of impaired safety signaling in Obsessive Compulsive Disorder. *Proceedings of the National Academy of Sciences*, *114*(12), 3216–3221. https://doi.org/10.1073/pnas.1609194114

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Attal, N., Masselin-Dubois, A., Martinez, V., Jayr, C., Albi, A., Fermanian, J., Bouhassira, D., & Baudic, S. (2014). Does cognitive functioning predict chronic pain? Results from a prospective surgical cohort. *Brain*, *137*(3), 904–917. https://doi.org/10.1093/brain/awt354

Austin, E. J. (2004). An investigation of the relationship between trait emotional intelligence and emotional task performance. *Personality and Individual Differences*, *36*(8), 1855–1864. https://doi.org/10.1016/j.paid.2003.07.006

Bach, P., Frischknecht, U., Bungert, M., Karl, D., Vollmert, C., Vollstädt-Klein, S., Lis, S., Kiefer, F., & Hermann, D. (2019). Effects of social exclusion and physical pain in chronic opioid maintenance treatment: fMRI correlates. *European Neuropsychopharmacology*, *29*(2), 291–305. https://doi.org/10.1016/j.euroneuro.2018.11.1109

Bach, P., Frischknecht, U., Reinhard, I., Bekier, N., Demirakca, T., Ende, G., Vollstädt-Klein, S., Kiefer, F., & Hermann, D. (2021). Impaired working memory performance in opioid-dependent patients is related to reduced insula gray matter volume: A voxel-based morphometric study. *European Archives of Psychiatry and Clinical Neuroscience*, *271*(5), 813–822. https://doi.org/10.1007/s00406-019-01052-7

Back, S. E., Gros, D. F., McCauley, J. L., Flanagan, J. C., Cox, E., Barth, K. S., & Brady, K. T. (2014). Laboratory-induced cue reactivity among individuals with prescription opioid dependence. *Addictive Behaviors*, *39*(8), 1217–1223. https://doi.org/10.1016/j.addbeh.2014.04.007

Badiani, A., Belin, D., Epstein, D., Calu, D., & Shaham, Y. (2011). Opiate versus psychostimulant addiction: The differences do matter. *Nature Reviews Neuroscience*, *12*(11), 685–700. https://doi.org/10.1038/nrn3104

Bagley, E. E., Chieng, B. C. H., Christie, M. J., & Connor, M. (2005). Opioid tolerance in periaqueductal gray neurons isolated from mice chronically treated with morphine. *British Journal of Pharmacology*, *146*(1), 68–76. https://doi.org/10.1038/sj.bjp.0706315

Baldacchino, A., Balfour, D. J. K., & Matthews, K. (2015). Impulsivity and opioid drugs: Differential effects of heroin, methadone and prescribed analgesic medication. *Psychological Medicine*, *45*(6), 1167–1179. https://doi.org/10.1017/S0033291714002189

Baldacchino, A., Balfour, D. J. K., Passetti, F., Humphris, G., & Matthews, K. (2012). Neuropsychological consequences of chronic opioid use: A quantitative review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *36*(9), 2056–2068. https://doi.org/10.1016/j.neubiorev.2012.06.006

Baldacchino, A., Tolomeo, S., Balfour, D. J., & Matthews, K. (2019). Profiles of visuospatial memory dysfunction in opioid-exposed and dependent populations. *Psychological Medicine*, *49*(07), 1174–1184. https://doi.org/10.1017/S0033291718003318

Baliki, M., Apkarian, A., Vigotsky, A., Rached, G., Jabakhanji, R., Huang, L., Branco, P., Barroso, J., Cong, O., Centeno, M. V., Cermak, S., Griffith, J., Schnitzer, T., & Wasan, A. (2024). *Multi-receptor whole-cortex adaptations with long-term opioid use in chronic pain*. In Review. https://doi.org/10.21203/rs.3.rs-5061297/v1

Ballantyne, J. C., & Shin, N. S. (2008). Efficacy of Opioids for Chronic Pain: A Review of the Evidence. *The Clinical Journal of Pain*, *24*(6), 469–478. https://doi.org/10.1097/AJP.0b013e31816b2f26

Ballantyne, J. C., Sullivan, M. D., & Koob, G. F. (2019). Refractory dependence on opioid analgesics. *Pain*, *160*(12), 2655–2660. https://doi.org/10.1097/j.pain.0000000000001680

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Basden, S. L., Orr, S. P., & Otto, M. W. (2016). Impaired De Novo Fear Conditioning in Opiate-Dependent Outpatients. *Cognitive Therapy and Research*, *40*(6), 824–830. https://doi.org/10.1007/s10608-016-9786-9

Becerra, L., Harter, K., Gonzalez, R. G., & Borsook, D. (2006). Functional Magnetic Resonance Imaging Measures of the Effects of Morphine on Central Nervous System Circuitry in Opioid-Naive Healthy Volunteers. *Anesthesia & Analgesia*, *103*(1), 208–216. https://doi.org/10.1213/01.ane.0000221457.71536.e0

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment*, *67*(3), 588–597. https://doi.org/10.1207/s15327752jpa6703_13

Becker, W. C., Sullivan, L. E., Tetrault, J. M., Desai, R. A., & Fiellin, D. A. (2008). Non-medical use, abuse and dependence on prescription opioids among U.S. adults: Psychiatric, medical and substance use correlates. *Drug and Alcohol Dependence*, *94*(1–3), 38–47. https://doi.org/10.1016/j.drugalcdep.2007.09.018

Belin, D., & Everitt, B. J. (2008). Cocaine Seeking Habits Depend upon Dopamine-Dependent Serial Connectivity Linking the Ventral with the Dorsal Striatum. *Neuron*, *57*(3), 432–441. https://doi.org/10.1016/j.neuron.2007.12.019

Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.2009.00972.x

Bernstein, D. P., Fink, L., Handelsman, L., & Foote, J. (2011). *Childhood Trauma Questionnaire* [Dataset]. https://doi.org/10.1037/t02080-000

Berridge, K. C. (2012). From prediction error to incentive salience: Mesolimbic computation of reward motivation. *European Journal of Neuroscience*, *35*(7), 1124–1143. https://doi.org/10.1111/j.1460-9568.2012.07990.x

Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, *71*(8), 670–679. https://doi.org/10.1037/amp0000059

Bershad, A. K., Jaffe, J. H., Childs, E., & De Wit, H. (2015). Opioid partial agonist buprenorphine dampens responses to psychosocial stress in humans. *Psychoneuroendocrinology*, *52*, 281–288. https://doi.org/10.1016/j.psyneuen.2014.12.004

Bershad, A. K., Miller, M. A., Norman, G. J., & De Wit, H. (2018). Effects of opioid- and non-opioid analgesics on responses to psychosocial stress in humans. *Hormones and Behavior*, *102*, 41–47. https://doi.org/10.1016/j.yhbeh.2018.04.009

Bershad, A. K., Seiden, J. A., & De Wit, H. (2016). Effects of buprenorphine on responses to social stimuli in healthy adults. *Psychoneuroendocrinology*, *63*, 43–49. https://doi.org/10.1016/j.psyneuen.2015.09.011

Besel, L. D. S., & Yuille, J. C. (2010). Individual differences in empathy: The role of facial expression recognition. *Personality and Individual Differences*, *49*(2), 107–112. https://doi.org/10.1016/j.paid.2010.03.013

Biernacki, K., Terrett, G., McLennan, S. N., Labuschagne, I., Morton, P., & Rendell, P. G. (2018). Decision-making, somatic markers and emotion processing in opiate users. *Psychopharmacology*, *235*(1), 223–232. https://doi.org/10.1007/s00213-017-4760-0

Bland, A. R., & Ersche, K. D. (2020). Deficits in recognizing female facial expressions related to social network in cocaine-addicted men. *Drug and Alcohol Dependence*, *216*, 108247. https://doi.org/10.1016/j.drugalcdep.2020.108247

Bland, A. R., Roiser, J. P., Mehta, M. A., Schei, T., Boland, H., Campbell-Meiklejohn, D. K., Emsley, R. A., Munafo, M. R., Penton-Voak, I. S., Seara-Cardoso, A., Viding, E., Voon, V., Sahakian, B. J., Robbins, T. W., & Elliott, R. (2016). EMOTICOM: A Neuropsychological Test Battery to Evaluate Emotion, Motivation, Impulsivity, and Social Cognition. *Frontiers in Behavioral Neuroscience*, *10*. https://doi.org/10.3389/fnbeh.2016.00025

Blanken, P., Hendriks, V. M., Koeter, M. W. J., Van Ree, J. M., & Van Den Brink, W. (2012). Craving and illicit heroin use among patients in heroin-assisted treatment. *Drug and Alcohol Dependence*, *120*(1–3), 74–80. https://doi.org/10.1016/j.drugalcdep.2011.06.025

Boehme, R., Deserno, L., Gleich, T., Katthagen, T., Pankow, A., Behr, J., Buchert, R., Roiser, J. P., Heinz, A., & Schlagenhauf, F. (2015). Aberrant Salience Is Related to Reduced Reinforcement Learning Signals and Elevated Dopamine Synthesis Capacity in Healthy Adults. *Journal of Neuroscience*, *35*(28), 10103–10111. https://doi.org/10.1523/JNEUROSCI.0805-15.2015

Bolívar, H. A., Klemperer, E. M., Coleman, S. R. M., DeSarno, M., Skelly, J. M., & Higgins, S. T. (2021). Contingency Management for Patients Receiving Medication for Opioid Use Disorder: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, *78*(10), 1092. https://doi.org/10.1001/jamapsychiatry.2021.1969

Bouton, M. E. (2021). Context, attention, and the switch between habit and goal-direction in behavior. *Learning & Behavior*, *49*(4), 349–362. https://doi.org/10.3758/s13420-021-00488-z

Brownstein, M. J. (1993). A brief history of opiates, opioid peptides, and opioid receptors. *Proceedings of the National Academy of Sciences*, *90*(12), 5391–5393. https://doi.org/10.1073/pnas.90.12.5391

Burroughs, W. S. (2008). *Junky: The definitive text of 'Junk'* (Repr.). Penguin.

Cahill, C. M., Walwyn, W., Taylor, A. M. W., Pradhan, A. A. A., & Evans, C. J. (2016). Allostatic Mechanisms of Opioid Tolerance Beyond Desensitization and Downregulation. *Trends in Pharmacological Sciences*, *37*(11), 963–976. https://doi.org/10.1016/j.tips.2016.08.002

Carlyle, M., Broomby, R., Simpson, G., Hannon, R., Fawaz, L., Mollaahmetoglu, O. M., Drain, J., Mostazir, M., & Morgan, C. J. A. (2021). A randomised, double-blind study investigating the relationship between early childhood trauma and the rewarding effects of morphine. *Addiction Biology*, *26*(6), e13047. https://doi.org/10.1111/adb.13047

Carlyle, M., Rowley, M., Stevens, T., Karl, A., & Morgan, C. J. A. (2020). Impaired empathy and increased anger following social exclusion in non-intoxicated opioid users. *Psychopharmacology*, *237*(2), 419–430. https://doi.org/10.1007/s00213-019-05378-x

Cavaiola, A. A., Fulmer, B. A., & Stout, D. (2015). The Impact of Social Support and Attachment Style on Quality of Life and Readiness to Change in a Sample of Individuals Receiving Medication-Assisted Treatment for Opioid Dependence. *Substance Abuse*, *36*(2), 183–191. https://doi.org/10.1080/08897077.2015.1019662

Ceceli, A. O., Huang, Y., Kronberg, G., Malaker, P., Miller, P., King, S. G., Gaudreault, P.-O., McClain, N., Gabay, L., Vasa, D., Newcorn, J. H., Ekin, D., Alia-Klein, N., & Goldstein, R. Z. (2023). Common and distinct fronto-striatal volumetric changes in heroin and cocaine use disorders. *Brain*, *146*(4), 1662–1671. https://doi.org/10.1093/brain/awac366

Ceceli, A. O., King, S. G., McClain, N., Alia-Klein, N., & Goldstein, R. Z. (2023). The Neural Signature of Impaired Inhibitory Control in Individuals with Heroin Use Disorder. *The Journal of Neuroscience*, *43*(1), 173–182. https://doi.org/10.1523/JNEUROSCI.1237-22.2022

Centers for Disease Control and Prevention. (2024a). *CDC WISQARS*. https://wisqars.cdc.gov/

Centers for Disease Control and Prevention. (2024b). *NCHS – Drug Poisoning Mortality by State: United States*. https://www.cdc.gov/nchs/nvss/drug-overdose-deaths.htm#find-our-data

Chang, V. N., & Peters, J. (2023). Neural circuits controlling choice behavior in opioid addiction. *Neuropharmacology*, *226*, 109407. https://doi.org/10.1016/j.neuropharm.2022.109407

Chelnokova, O., Laeng, B., Løseth, G., Eikemo, M., Willoch, F., & Leknes, S. (2016). The μ-opioid system promotes visual attention to faces and eyes. *Social Cognitive and Affective Neuroscience*, *11*(12), 1902–1909. https://doi.org/10.1093/scan/nsw116

Chou, R. (2015). Long-Term Opioid Therapy for Chronic Pain. *Annals of Internal Medicine*, *163*(2), 148. https://doi.org/10.7326/L15-5109-3

Christie, M. J. (2008). Cellular neuroadaptations to chronic opioids: Tolerance, withdrawal and addiction. *British Journal of Pharmacology*, *154*(2), 384–396. https://doi.org/10.1038/bjp.2008.100

Christie, N. C. (2021). The role of social isolation in opioid addiction. *Social Cognitive and Affective Neuroscience*, *16*(7), 645–656. https://doi.org/10.1093/scan/nsab029

Cicero, T. J., Ellis, M. S., Surratt, H. L., & Kurtz, S. P. (2014). The Changing Face of Heroin Use in the United States: A Retrospective Analysis of the Past 50 Years. *JAMA Psychiatry*, *71*(7), 821. https://doi.org/10.1001/jamapsychiatry.2014.366

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed). Taylor and Francis.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385. https://doi.org/10.2307/2136404

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x

Comer, S. D., Sullivan, M. A., Vosburg, S. K., Kowalczyk, W. J., & Houser, J. (2010). Abuse liability of oxycodone as a function of pain and drug use history. *Drug and Alcohol Dependence*, *109*(1–3), 130–138. https://doi.org/10.1016/j.drugalcdep.2009.12.018

Compton, W. M., Stinson, F. S., & Grant, B. F. (2005). *Prevalence, Correlates, and Comorbidity of DSM-IV Antisocial Personality Syndromes and Alcohol and Specific Drug Use Disorders in the United States: Results From the National Epidemiologic Survey on Alcohol and Related Conditions*.

Corden, B., Critchley, H. D., Skuse, D., & Dolan, R. J. (2006). Fear Recognition Ability Predicts Differences in Social Cognitive and Neural Functioning in Men. *Journal of Cognitive Neuroscience*, *18*(6), 889–897. https://doi.org/10.1162/jocn.2006.18.6.889

Craparo, G., Gori, A., Dell'Aera, S., Costanzo, G., Fasciano, S., Tomasello, A., & Vicario, C. M. (2016). Impaired emotion recognition is linked to alexithymia in heroin addicts. *PeerJ*, *4*, e1864. https://doi.org/10.7717/peerj.1864

Critchley, H. D. (2002). Review: Electrodermal Responses: What Happens in the Brain. *The Neuroscientist*, *8*(2), 132–142. https://doi.org/10.1177/107385840200800209

Crockett, M., & Cools, R. (2015). Serotonin and aversive processing in affective and social decision-making. *Current Opinion in Behavioral Sciences*, *5*, 64–70. https://doi.org/10.1016/j.cobeha.2015.08.005

Daglish, M. R. C., Williams, T. M., Wilson, S. J., Taylor, L. G., Eap, C. B., Augsburger, M., Giroud, C., Brooks, D. J., Myles, J. S., Grasby, P., Lingford-Hughes, A. R., & Nutt, D. J. (2008). Brain dopamine response in human opioid addiction. *British Journal of Psychiatry*, *193*(1), 65–72. https://doi.org/10.1192/bjp.bp.107.041228

Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience & Biobehavioral Reviews*, *36*(10), 2288–2304. https://doi.org/10.1016/j.neubiorev.2012.08.006

De Vries, F., Bruin, M., Lobatto, D. J., Dekkers, O. M., Schoones, J. W., Van Furth, W. R., Pereira, A. M., Karavitaki, N., Biermasz, N. R., & Zamanipoor Najafabadi, A. H. (2020). Opioids and Their Endocrine Effects: A Systematic Review and Meta-analysis. *The Journal of Clinical Endocrinology & Metabolism*, *105*(4), 1020–1029. https://doi.org/10.1210/clinem/dgz022

De Wit, S., Corlett, P. R., Aitken, M. R., Dickinson, A., & Fletcher, P. C. (2009). Differential Engagement of the Ventromedial Prefrontal Cortex by Goal-Directed and Habitual Behavior toward Food Pictures in Humans. *The Journal of Neuroscience*, *29*(36), 11330–11338. https://doi.org/10.1523/JNEUROSCI.1639-09.2009

De Wit, S., & Dickinson, A. (2009). Associative theories of goal-directed behaviour: A case for animal–human translational models. *Psychological Research Psychologische Forschung*, *73*(4), 463–476. https://doi.org/10.1007/s00426-009-0230-6

De Wit, S., Standing, H. R., DeVito, E. E., Robinson, O. J., Ridderinkhof, K. R., Robbins, T. W., & Sahakian, B. J. (2012). Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. *Psychopharmacology*, *219*(2), 621–631. https://doi.org/10.1007/s00213-011-2563-2

De Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., Van De Vijver, I., & Ridderinkhof, K. R. (2012). Corticostriatal Connectivity Underlies Individual Differences in the Balance between Habitual and Goal-Directed Action Control. *The Journal of Neuroscience*, *32*(35), 12066–12075. https://doi.org/10.1523/JNEUROSCI.1088-12.2012

Dejoie, J. M., Senia, N., Konova, A. B., Smith, D. V., & Fareri, D. S. (2024). Common and distinct drug cue reactivity patterns associated with cocaine and heroin: An fMRI meta-analysis. *Imaging Neuroscience*, *2*, 1–18. https://doi.org/10.1162/imag_a_00211

Delgado, M. R., Nearing, K. I., LeDoux, J. E., & Phelps, E. A. (2008). Neural Circuitry Underlying the Regulation of Conditioned Fear and Its Relation to Extinction. *Neuron*, *59*(5), 829–838. https://doi.org/10.1016/j.neuron.2008.06.029

Derksen, M., & Morawski, J. (2022). Kinds of Replication: Examining the Meanings of "Conceptual Replication" and "Direct Replication". *Perspectives on Psychological Science*, *17*(5), 1490–1505. https://doi.org/10.1177/17456916211041116

Deroche-Gamonet, V., Belin, D., & Piazza, P. V. (2004). Evidence for Addiction-like Behavior in the Rat. *Science*, *305*(5686), 1014–1017. https://doi.org/10.1126/science.1099020

Diano, M., Tamietto, M., Celeghin, A., Weiskrantz, L., Tatu, M.-K., Bagnis, A., Duca, S., Geminiani, G., Cauda, F., & Costa, T. (2017). Dynamic Changes in Amygdala Psychophysiological Connectivity Reveal Distinct Neural Networks for Facial Expressions of Basic Emotions. *Scientific Reports*, *7*(1), 45260. https://doi.org/10.1038/srep45260

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *308*(1135), 67–78. https://doi.org/10.1098/rstb.1985.0010

Dickinson, A., Wood, N., & Smith, J. W. (2002). Alcohol Seeking by Rats: Action or Habit? *The Quarterly Journal of Experimental Psychology Section B*, *55*(4b), 331–348. https://doi.org/10.1080/0272499024400016

DiFranza, J. R. (2016). Can tobacco dependence provide insights into other drug addictions? *BMC Psychiatry*, *16*(1), 365. https://doi.org/10.1186/s12888-016-1074-4

Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences*, *1*, 101–106. https://doi.org/10.1016/j.cobeha.2014.10.007

Doverty, M., White, J. M., Somogyi, A. A., Bochner, F., Ali, R., & Ling, W. (2001). Hyperalgesic responses in methadone maintenance patients. *Pain*, *90*(1), 91–96. https://doi.org/10.1016/S0304-3959(00)00391-2

Dugosh, K., Abraham, A., Seymour, B., McLoyd, K., Chalk, M., & Festinger, D. (2016). A Systematic Review on the Use of Psychosocial Interventions in Conjunction With Medications for the Treatment of Opioid Addiction. *Journal of Addiction Medicine*, *10*(2), 93–103. https://doi.org/10.1097/ADM.0000000000000193

Eikemo, M., Løseth, G. E., & Leknes, S. (2021). Do endogenous opioids mediate or fine-tune human pain relief? *Pain*, *162*(12), 2789–2791. https://doi.org/10.1097/j.pain.0000000000002286

Eikemo, M., Meier, I. M., Løseth, G. E., Trøstheim, M., Ørstavik, N., Jensen, E. N., Garland, E. L., Berna, C., Ernst, G., & Leknes, S. (2023). Opioid analgesic effects on subjective well-being in the operating theatre*. *Anaesthesia*, *78*(9), 1102–1111. https://doi.org/10.1111/anae.16069

Eippert, F., Bingel, U., Schoell, E., Yacubian, J., & Büchel, C. (2008). Blockade of Endogenous Opioid Neurotransmission Enhances Acquisition of Conditioned Fear in Humans. *The Journal of Neuroscience*, *28*(21), 5465–5472. https://doi.org/10.1523/JNEUROSCI.5336-07.2008

Eldar, E., Hauser, T. U., Dayan, P., & Dolan, R. J. (2016). Striatal structure and function predict individual biases in learning to avoid pain. *Proceedings of the National Academy of Sciences*, *113*(17), 4812–4817. https://doi.org/10.1073/pnas.1519829113

Elliott, K. R., & Jones, E. (2019). The association between frequency of opioid misuse and opioid use disorder among youth and adults in the United States. *Drug and Alcohol Dependence*, *197*, 73–77. https://doi.org/10.1016/j.drugalcdep.2019.01.008

Ersche, K. D., Gillan, C. M., Jones, P. S., Williams, G. B., Ward, L. H. E., Luijten, M., De Wit, S., Sahakian, B. J., Bullmore, E. T., & Robbins, T. W. (2016). Carrots and sticks fail to change behavior in cocaine addiction. *Science*, *352*(6292), 1468–1471. https://doi.org/10.1126/science.aaf3700

Ersche, K. D., Hagan, C. C., Smith, D. G., Jones, P. S., Calder, A. J., & Williams, G. B. (2015). In the face of threat: Neural and endocrine correlates of impaired facial emotion recognition in cocaine dependence. *Translational Psychiatry*, *5*(5), e570–e570. https://doi.org/10.1038/tp.2015.58

Ersche, K. D., Lim, T. V., Murley, A. G., Rua, C., Vaghi, M. M., White, T. L., Williams, G. B., & Robbins, T. W. (2021). Reduced Glutamate Turnover in the Putamen Is Linked With Automatic Habits in Human Cocaine Addiction. *Biological Psychiatry*, *89*(10), 970–979. https://doi.org/10.1016/j.biopsych.2020.12.009

European Monitoring Centre for Drugs and Drug Addiction. (2023). *Drug-induced deaths – the current situation in Europe (European Drug Report 2023)*. https://www.euda.europa.eu/publications/european-drug-report/2023/drug-induced-deaths_en

Everitt, B. J., Belin, D., Economidou, D., Pelloux, Y., Dalley, J. W., & Robbins, T. W. (2008). Neural mechanisms underlying the vulnerability to develop compulsive drug-seeking habits and addiction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1507), 3125–3135. https://doi.org/10.1098/rstb.2008.0089

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, *8*(11), 1481–1489. https://doi.org/10.1038/nn1579

Everitt, B. J., & Robbins, T. W. (2016). Drug Addiction: Updating Actions to Habits to Compulsions Ten Years On. *Annual Review of Psychology*, *67*(1), 23–50. https://doi.org/10.1146/annurev-psych-122414-033457

Fanselow, M. S., & Bolles, R. C. (1979). Naloxone and shock-elicited freezing in the rat. *Journal of Comparative and Physiological Psychology*, *93*(4), 736–744. https://doi.org/10.1037/h0077609

Fatseas, M., Denis, C., Massida, Z., Verger, M., Franques-Rénéric, P., & Auriacombe, M. (2011). Cue-Induced Reactivity, Cortisol Response and Substance Use Outcome in Treated Heroin Dependent Individuals. *Biological Psychiatry*, *70*(8), 720–727. https://doi.org/10.1016/j.biopsych.2011.05.015

Fazeli, S., & Büchel, C. (2018). Pain-Related Expectation and Prediction Error Signals in the Anterior Insula Are Not Related to Aversiveness. *The Journal of Neuroscience*, *38*(29), 6461–6474. https://doi.org/10.1523/JNEUROSCI.0671-18.2018

Fields, H. L., & Margolis, E. B. (2015). Understanding opioid reward. *Trends in Neurosciences*, *38*(4), 217–225. https://doi.org/10.1016/j.tins.2015.01.002

Franken, I. H. A., Hendriks, V. M., & Van Den Brink, W. (2002). Initial validation of two opiate craving questionnaires. *Addictive Behaviors*, *27*(5), 675–685. https://doi.org/10.1016/S0306-4603(01)00201-5

Fullana, M. A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., Radua, J., & Harrison, B. J. (2018). Fear extinction in the human brain: A meta-analysis of fMRI studies in healthy participants. *Neuroscience & Biobehavioral Reviews*, *88*, 16–25. https://doi.org/10.1016/j.neubiorev.2018.03.002

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

Gamer, M., & Buchel, C. (2009). Amygdala Activation Predicts Gaze toward Fearful Eyes. *Journal of Neuroscience*, *29*(28), 9123–9126. https://doi.org/10.1523/JNEUROSCI.1883-09.2009

Garbusow, M., Ebrahimi, C., Riemerschmid, C., Daldrup, L., Rothkirch, M., Chen, K., Chen, H., Belanger, M. J., Hentschel, A., Smolka, M. N., Heinz, A., Pilhatsch, M., & Rapp, M. A. (2022). Pavlovian-to-Instrumental Transfer across Mental Disorders: A Review. *Neuropsychobiology*, *81*(5), 418–437. https://doi.org/10.1159/000525579

Garbusow, M., Schad, D. J., Sebold, M., Friedel, E., Bernhardt, N., Koch, S. P., Steinacher, B., Kathmann, N., Geurts, D. E. M., Sommer, C., Müller, D. K., Nebe, S., Paul, S., Wittchen, H.-U., Zimmermann, U. S., Walter, H., Smolka, M. N., Sterzer, P., Rapp, M. A., … Heinz, A. (2016). Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence: PIT and alcohol relapse. *Addiction Biology*, *21*(3), 719–731. https://doi.org/10.1111/adb.12243

Garland, E. L. (2024). Mindfulness-Oriented Recovery Enhancement: Implementing an evidence-based intervention for chronic pain, opioid use, and opioid addiction in clinical settings. *British Journal of Clinical Pharmacology*, *90*(12), 3028–3035. https://doi.org/10.1111/bcp.16147

Garland, E. L., Atchley, R. M., Hanley, A. W., Zubieta, J.-K., & Froeliger, B. (2019). Mindfulness-Oriented Recovery Enhancement remediates hedonic dysregulation in opioid users: Neural and affective evidence of target engagement. *Science Advances*, *5*(10), eaax1569. https://doi.org/10.1126/sciadv.aax1569

Garland, E. L., Froeliger, B. E., Passik, S. D., & Howard, M. O. (2013). Attentional bias for prescription opioid cues among opioid dependent chronic pain patients. *Journal of Behavioral Medicine*, *36*(6), 611–620. https://doi.org/10.1007/s10865-012-9455-8

Garland, E. L., & Howard, M. O. (2014). Opioid attentional bias and cue-elicited craving predict future risk of prescription opioid misuse among chronic pain patients. *Drug and Alcohol Dependence*, *144*, 283–287. https://doi.org/10.1016/j.drugalcdep.2014.09.014

Garland, E. L., Manusov, E. G., Froeliger, B., Kelly, A., Williams, J. M., & Howard, M. O. (2014). Mindfulness-oriented recovery enhancement for chronic pain and prescription opioid misuse: Results from an early-stage randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *82*(3), 448–459. https://doi.org/10.1037/a0035798

Garofalo, S., & Robbins, T. W. (2017). Triggering Avoidance: Dissociable Influences of Aversive Pavlovian Conditioned Stimuli on Human Instrumental Behavior. *Frontiers in Behavioral Neuroscience*, *11*. https://doi.org/10.3389/fnbeh.2017.00063

Gaser, C., Dahnke, R., Thompson, P. M., Kurth, F., Luders, E., & the Alzheimer's Disease Neuroimaging Initiative. (2024). CAT: A computational anatomy toolbox for the analysis of structural MRI data. *GigaScience*, *13*, giae049. https://doi.org/10.1093/gigascience/giae049

German Federal Ministry of Health. (2018). *Schätzung Opioidabhängiger in Deutschland—Estimation of opioid-addicted patients in Germany*. www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Drogen_und_Sucht/Berichte/AbschlussberichtOpiS-Bericht_150518.pdf

Gerra, G., Baldaro, B., Zaimovic, A., Moi, G., Bussandri, M., Raggi, M. A., & Brambilla, F. (2003). Neuroendocrine responses to experimentally-induced emotions among abstinent opioid-dependent subjects. *Drug and Alcohol Dependence*, *71*(1), 25–35. https://doi.org/10.1016/S0376-8716(03)00065-6

Geurts, D. E. M., Huys, Q. J. M., Den Ouden, H. E. M., & Cools, R. (2013). Aversive Pavlovian Control of Instrumental Behavior in Humans. *Journal of Cognitive Neuroscience*, *25*(9), 1428–1441. https://doi.org/10.1162/jocn_a_00425

Gibson, A., Degenhardt, L., Mattick, R. P., Ali, R., White, J., & O'Brien, S. (2008). Exposure to opioid maintenance treatment reduces long-term mortality. *Addiction*, *103*(3), 462–468. https://doi.org/10.1111/j.1360-0443.2007.02090.x

Gillan, C. M., Apergis-Schoute, A. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W. (2015). Functional Neuroimaging of Avoidance Habits in Obsessive-Compulsive Disorder. *American Journal of Psychiatry*, *172*(3), 284–293. https://doi.org/10.1176/appi.ajp.2014.14040525

Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W. (2014). Enhanced Avoidance Habits in Obsessive-Compulsive Disorder. *Biological Psychiatry*, *75*(8), 631–638. https://doi.org/10.1016/j.biopsych.2013.02.002

Gläscher, J., & Gitelman, D. (2008). *Contrast weights in flexible factorial design with multiple groups of subjects*. SPM@JISCMAIL.AC.UK.

Glover, G. H. (2011). Overview of Functional Magnetic Resonance Imaging. *Neurosurgery Clinics of North America*, *22*(2), 133–139. https://doi.org/10.1016/j.nec.2010.11.001

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: Neuroimaging findings and clinical implications. *Nature Reviews Neuroscience*, *12*(11), 652–669. https://doi.org/10.1038/nrn3119

Good, A. J., & Westbrook, R. F. (1995). Effects of a microinjection of morphine into the amygdala on the acquisition and expression of conditioned fear and hypoalgesia in rats. *Behavioral Neuroscience*, *109*(4), 631–641. https://doi.org/10.1037/0735-7044.109.4.631

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

Gossop, M. (1990). The development of a short opiate withdrawal scale (SOWS). *Addictive Behaviors*, *15*(5), 487–490. https://doi.org/10.1016/0306-4603(90)90036-W

Gradin, V. B., Baldacchino, A., Balfour, D., Matthews, K., & Steele, J. D. (2014). Abnormal Brain Activity During a Reward and Loss Task in Opiate-Dependent Patients Receiving Methadone Maintenance Therapy. *Neuropsychopharmacology*, *39*(4), 885–894. https://doi.org/10.1038/npp.2013.289

Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications*, *4*(1), 2264. https://doi.org/10.1038/ncomms3264

Griffin, M. L., Bennett, H. E., Fitzmaurice, G. M., Hill, K. P., Provost, S. E., & Weiss, R. D. (2015). Health-related quality of life among prescription opioid-dependent patients: Results from a multi-site study: Quality of life in prescription opioid dependence. *The American Journal on Addictions*, *24*(4), 308–314. https://doi.org/10.1111/ajad.12188

Gu, C., Li, P., Hu, B., Ouyang, X., Fu, J., Gao, J., Song, Z., Han, L., Ma, Y., Tian, S., & Hu, X. (2008). Chronic Morphine Selectively Impairs Cued Fear Extinction in Rats: Implications for Anxiety Disorders Associated with Opiate Use. *Neuropsychopharmacology*, *33*(3), 666–673. https://doi.org/10.1038/sj.npp.1301441

Guitart-Masip, M., Fuentemilla, L., Bach, D. R., Huys, Q. J. M., Dayan, P., Dolan, R. J., & Duzel, E. (2011). Action Dominates Valence in Anticipatory Representations in the Human Striatum and Dopaminergic Midbrain. *The Journal of Neuroscience*, *31*(21), 7867–7875. https://doi.org/10.1523/JNEUROSCI.6376-10.2011

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage*, *62*(1), 154–166. https://doi.org/10.1016/j.neuroimage.2012.04.024

Haasen, C., Verthein, U., Degkwitz, P., Berger, J., Krausz, M., & Naber, D. (2007). Heroin-assisted treatment for opioid dependence: Randomised controlled trial. *British Journal of Psychiatry*, *191*(1), 55–62. https://doi.org/10.1192/bjp.bp.106.026112

Han, B., Compton, W. M., Blanco, C., Crane, E., Lee, J., & Jones, C. M. (2017). Prescription Opioid Use, Misuse, and Use Disorders in U.S. Adults: 2015 National Survey on Drug Use and Health. *Annals of Internal Medicine*, *167*(5), 293. https://doi.org/10.7326/M17-0865

Harrison, B. J., Fullana, M. A., Via, E., Soriano-Mas, C., Vervliet, B., Martínez-Zalacaín, I., Pujol, J., Davey, C. G., Kircher, T., Straube, B., & Cardoner, N. (2017). Human ventromedial prefrontal cortex and the positive affective processing of safety signals. *NeuroImage*, *152*, 12–18. https://doi.org/10.1016/j.neuroimage.2017.02.080

Hay, J. L., White, J. M., Bochner, F., Somogyi, A. A., Semple, T. J., & Rounsefell, B. (2009). Hyperalgesia in Opioid-Managed Chronic Pain and Opioid-Dependent Patients. *The Journal of Pain*, *10*(3), 316–322. https://doi.org/10.1016/j.jpain.2008.10.003

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K.-O. (1991). The Fagerstrom Test for Nicotine Dependence: A revision of the Fagerstrom Tolerance Questionnaire. *Addiction*, *86*(9), 1119–1127. https://doi.org/10.1111/j.1360-0443.1991.tb01879.x

Hebart, M. N., & Gläscher, J. (2015). Serotonin and dopamine differentially affect appetitive and aversive general Pavlovian-to-instrumental transfer. *Psychopharmacology*, *232*(2), 437–451. https://doi.org/10.1007/s00213-014-3682-3

Heilig, M., Epstein, D. H., Nader, M. A., & Shaham, Y. (2016). Time to connect: Bringing social context into addiction neuroscience. *Nature Reviews Neuroscience*, *17*(9), 592–599. https://doi.org/10.1038/nrn.2016.67

Henschel, A. V., Flanagan, J. C., Augur, I. F., Jeffirs, S. M., & Back, S. E. (2022). Motives for prescription opioid use: The role of alexithymia and distress tolerance. *The American Journal on Addictions*, *31*(1), 55–60. https://doi.org/10.1111/ajad.13230

Heyes, C., & Dickinson, A. (1990). The Intentionality of Animal Action. *Mind & Language*, *5*(1), 87–103. https://doi.org/10.1111/j.1468-0017.1990.tb00154.x

Higgins, C., Smith, B. H., & Matthews, K. (2018). Incidence of iatrogenic opioid dependence or abuse in patients with pain who were exposed to opioid analgesic therapy: A systematic review and meta-analysis. *British Journal of Anaesthesia*, *120*(6), 1335–1344. https://doi.org/10.1016/j.bja.2018.03.009

Higgins, C., Smith, B. H., & Matthews, K. (2019). Evidence of opioid-induced hyperalgesia in clinical populations after chronic opioid exposure: A systematic review and meta-analysis. *British Journal of Anaesthesia*, *122*(6), e114–e126. https://doi.org/10.1016/j.bja.2018.09.019

Hogarth, L., Dickinson, A., & Duka, T. (2010). The associative basis of cue-elicited drug taking in humans. *Psychopharmacology*, *208*(3), 337–351. https://doi.org/10.1007/s00213-009-1735-9

Holmes, N. M., Marchand, A. R., & Coutureau, E. (2010). Pavlovian to instrumental transfer: A neurobehavioural perspective. *Neuroscience & Biobehavioral Reviews*, *34*(8), 1277–1295. https://doi.org/10.1016/j.neubiorev.2010.03.007

Honkalampi, K., Jokela, M., Lehto, S. M., Kivimäki, M., & Virtanen, M. (2022). Association between alexithymia and substance use: A systematic review and meta-analysis. *Scandinavian Journal of Psychology*, *63*(5), 427–438. https://doi.org/10.1111/sjop.12821

Hoskin, R., & Talmi, D. (2023). Adaptive coding of pain prediction error in the anterior insula. *European Journal of Pain*, *27*(6), 766–778. https://doi.org/10.1002/ejp.2093

Hser, Y.-I., Evans, E., Grella, C., Ling, W., & Anglin, D. (2015). Long-Term Course of Opioid Addiction. *Harvard Review of Psychiatry*, *23*(2), 76–89. https://doi.org/10.1097/HRP.0000000000000052

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. *PLoS Computational Biology*, *7*(4), e1002028. https://doi.org/10.1371/journal.pcbi.1002028

Hyman, S. M., Fox, H., Hong, K.-I. A., Doebrick, C., & Sinha, R. (2007). Stress and drug-cue-induced craving in opioid-dependent individuals in naltrexone treatment. *Experimental and Clinical Psychopharmacology*, *15*(2), 134–143. https://doi.org/10.1037/1064-1297.15.2.134

Icenhour, A., Langhorst, J., Benson, S., Schlamann, M., Hampel, S., Engler, H., Forsting, M., & Elsenbruch, S. (2015). Neural circuitry of abdominal pain-related fear learning and reinstatement in irritable bowel syndrome. *Neurogastroenterology & Motility*, *27*(1), 114–127. https://doi.org/10.1111/nmo.12489

Ipser, J. C., Terburg, D., Syal, S., Phillips, N., Solms, M., Panksepp, J., Malcolm-Smith, S., Thomas, K., Stein, D. J., & Van Honk, J. (2013). Reduced fear-recognition sensitivity following acute buprenorphine administration in healthy volunteers. *Psychoneuroendocrinology*, *38*(1), 166–170. https://doi.org/10.1016/j.psyneuen.2012.05.002

Jenewein, J., Moergeli, H., Sprott, H., Honegger, D., Brunner, L., Ettlin, D., Grillon, C., Bloch, K., Brügger, M., Schwegler, K., Schumacher, S., & Hasler, G. (2013). Fear-learning deficits in subjects with fibromyalgia syndrome? *European Journal of Pain*, *17*(9), 1374–1384. https://doi.org/10.1002/j.1532-2149.2013.00300.x

Jepma, M., Roy, M., Ramlakhan, K., Van Velzen, M., & Dahan, A. (2022). Different brain systems support learning from received and avoided pain during human pain-avoidance learning. *eLife*, *11*, e74149. https://doi.org/10.7554/eLife.74149

Ju, C., Wei, L., Man, K. K. C., Wang, Z., Ma, T.-T., Chan, A. Y. L., Brauer, R., Chui, C. S. L., Chan, E. W., Jani, Y. H., Hsia, Y., Wong, I. C. K., & Lau, W. C. Y. (2022). Global, regional, and national trends in opioid analgesic consumption from 2015 to 2019: A longitudinal study. *The Lancet Public Health*, *7*(4), e335–e346. https://doi.org/10.1016/S2468-2667(22)00013-5

Just, J. M., Bingener, L., Bleckwenn, M., Schnakenberg, R., & Weckbecker, K. (2018). Risk of opioid misuse in chronic non-cancer pain in primary care patients—A cross sectional study. *BMC Family Practice*, *19*(1), 92. https://doi.org/10.1186/s12875-018-0775-9

Kanen, J. W., Apergis-Schoute, A. M., Yellowlees, R., Arntz, F. E., Van Der Flier, F. E., Price, A., Cardinal, R. N., Christmas, D. M., Clark, L., Sahakian, B. J., Crockett, M. J., & Robbins, T. W. (2021). Serotonin depletion impairs both Pavlovian and instrumental reversal learning in healthy humans. *Molecular Psychiatry*, *26*(12), 7200–7210. https://doi.org/10.1038/s41380-021-01240-9

Kantonen, T., Karjalainen, T., Isojärvi, J., Nuutila, P., Tuisku, J., Rinne, J., Hietala, J., Kaasinen, V., Kalliokoski, K., Scheinin, H., Hirvonen, J., Vehtari, A., & Nummenmaa, L. (2020). Interindividual variability and lateralization of μ-opioid receptors in the human brain. *NeuroImage*, *217*, 116922. https://doi.org/10.1016/j.neuroimage.2020.116922

Kendler, K. S., Lönn, S. L., Ektor-Andersen, J., Sundquist, J., & Sundquist, K. (2023). Risk factors for the development of opioid use disorder after first opioid prescription: A Swedish national study. *Psychological Medicine*, *53*(13), 6223–6231. https://doi.org/10.1017/S003329172200349X

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is Avoiding an Aversive Outcome Rewarding? Neural Substrates of Avoidance Learning in the Human Brain. *PLoS Biology*, *4*(8), e233. https://doi.org/10.1371/journal.pbio.0040233

Kim, Y.-T., Kwon, D.-H., & Chang, Y. (2011). Impairments of facial emotion recognition and theory of mind in methamphetamine abusers. *Psychiatry Research*, *186*(1), 80–84. https://doi.org/10.1016/j.psychres.2010.06.027

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General*, *128*(1), 78–87. https://doi.org/10.1037/0096-3445.128.1.78

Kling, M. A., Carson, R. E., Borg, L., Zametkin, A., Matochik, J. A., Schluger, J., Herscovitch, P., Rice, K. C., Ho, A., Eckelman, W. C., & Kreek, M. J. (2000). Opioid receptor imaging with positron emission tomography and [(18)F]cyclofoxy in long-term, methadone-treated former heroin addicts. *Journal of Pharmacology and Experimental Therapeutics*, *295(3)*, 1070–1076.

Knisely, J. S., Wunsch, M. J., Cropsey, K. L., & Campbell, E. D. (2008). Prescription Opioid Misuse Index: A brief questionnaire to assess misuse. *Journal of Substance Abuse Treatment*, *35*(4), 380–386. https://doi.org/10.1016/j.jsat.2008.02.001

Koob, G. F. (2013). Negative reinforcement in drug addiction: The darkness within. *Current Opinion in Neurobiology*, *23*(4), 559–563. https://doi.org/10.1016/j.conb.2013.03.011

Koob, G. F. (2020). Neurobiology of Opioid Addiction: Opponent Process, Hyperkatifeia, and Negative Reinforcement. *Biological Psychiatry*, *87*(1), 44–53. https://doi.org/10.1016/j.biopsych.2019.05.023

Koob, G. F., & Volkow, N. D. (2016). Neurobiology of addiction: A neurocircuitry analysis. *The Lancet Psychiatry*, *3*(8), 760–773. https://doi.org/10.1016/S2215-0366(16)00104-8

Kornreich, C., Foisy, M.-L., Philippot, P., Dan, B., Tecco, J., Noël, X., Hess, U., Pelc, I., & Verbanck, P. (2003). Impaired emotional facial expression recognition in alcoholics, opiate dependence subjects, methadone maintained subjects and mixed alcohol-opiate antecedents subjects compared with normal controls. *Psychiatry Research*, *119*(3), 251–260. https://doi.org/10.1016/S0165-1781(03)00130-6

Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, *10*(11), 1437–1448. https://doi.org/10.1093/scan/nsv032

Kroll, S. L., Williams, D. P., Thoma, M., Staib, M., Binz, T. M., Baumgartner, M. R., Kirschbaum, C., Thayer, J. F., & Quednow, B. B. (2019). Non-medical prescription opioid users exhibit dysfunctional physiological stress responses to social rejection. *Psychoneuroendocrinology*, *100*, 264–275. https://doi.org/10.1016/j.psyneuen.2018.09.023

Krypotos, A.-M. (2015). Avoidance learning: A review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00189

Kuhn, M., Gerlicher, A. M. V., & Lonsdorf, T. B. (2022). Navigating the manyverse of skin conductance response quantification approaches – A direct comparison of TROUGH-TO-PEAK , baseline correction, and model-based approaches in Ledalab and PSPM. *Psychophysiology*, *59*(9), e14058. https://doi.org/10.1111/psyp.14058

Kumar, N., Oles, W., Howell, B. A., Janmohamed, K., Lee, S. T., Funaro, M. C., O'Connor, P. G., & Alexander, M. (2021). The role of social network support in treatment outcomes for medication for opioid use disorder: A systematic review. *Journal of Substance Abuse Treatment*, *127*, 108367. https://doi.org/10.1016/j.jsat.2021.108367

Laing, P. A. F., & Harrison, B. J. (2021). Safety learning and the Pavlovian conditioned inhibition of fear in humans: Current state and future directions. *Neuroscience & Biobehavioral Reviews*, *127*, 659–674. https://doi.org/10.1016/j.neubiorev.2021.05.014

Lee, M. C., Wanigasekera, V., & Tracey, I. (2014). Imaging opioid analgesia in the human brain and its potential relevance for understanding opioid use in chronic pain. *Neuropharmacology*, *84*, 123–130. https://doi.org/10.1016/j.neuropharm.2013.06.035

Leppä, M., Korvenoja, A., Carlson, S., Timonen, P., Martinkauppi, S., Ahonen, J., Rosenberg, P. H., Aronen, H. J., & Kalso, E. (2006). Acute opioid effects on human brain as revealed by functional magnetic resonance imaging. *NeuroImage*, *31*(2), 661–669. https://doi.org/10.1016/j.neuroimage.2005.12.019

Lewis, A. H., Niznikiewicz, M. A., Delamater, A. R., & Delgado, M. R. (2013). Avoidance-based human Pavlovian-to-instrumental transfer. *European Journal of Neuroscience*, *38*(12), 3740–3748. https://doi.org/10.1111/ejn.12377

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience*, *14*(10), 1250–1252. https://doi.org/10.1038/nn.2904

Li, Q., Li, W., Wang, H., Wang, Y., Zhang, Y., Zhu, J., Zheng, Y., Zhang, D., Wang, L., Li, Y., Yan, X., Chang, H., Fan, M., Li, Z., Tian, J., Gold, M. S., Wang, W., & Liu, Y. (2015). Predicting subsequent relapse by drug-related cue-induced brain activation in heroin addiction: An event-related functional magnetic resonance imaging study. *Addiction Biology*, *20*(5), 968–978. https://doi.org/10.1111/adb.12182

Li, S. S. Y., & McNally, G. P. (2014). The conditions that promote fear learning: Prediction error and Pavlovian fear conditioning. *Neurobiology of Learning and Memory*, *108*, 14–21. https://doi.org/10.1016/j.nlm.2013.05.002

Li, Y., Simmler, L. D., Van Zessen, R., Flakowski, J., Wan, J.-X., Deng, F., Li, Y.-L., Nautiyal, K. M., Pascoli, V., & Lüscher, C. (2021). Synaptic mechanism underlying serotonin modulation of transition to cocaine addiction. *Science*, *373*(6560), 1252–1256. https://doi.org/10.1126/science.abi9086

Lim, T. V., Cardinal, R. N., Savulich, G., Jones, P. S., Moustafa, A. A., Robbins, T. W., & Ersche, K. D. (2019). Impairments in reinforcement learning do not explain enhanced habit formation in cocaine use disorder. *Psychopharmacology*, *236*(8), 2359–2371. https://doi.org/10.1007/s00213-019-05330-z

Lin, J. C., Chu, L. F., Stringer, E. A., Baker, K. S., Sayyid, Z. N., Sun, J., Campbell, K. A., & Younger, J. W. (2016). One Month of Oral Morphine Decreases Gray Matter Volume in the Right Amygdala of Individuals with Low Back Pain: Confirmation of Previously Reported Magnetic Resonance Imaging Results. *Pain Medicine*, *17*(8), 1497–1504. https://doi.org/10.1093/pm/pnv047

Logothetis, N. K. (2003). The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal. *The Journal of Neuroscience*, *23*(10), 3963–3971. https://doi.org/10.1523/JNEUROSCI.23-10-03963.2003

Lonsdorf, T. B., Menz, M. M., Andreatta, M., Fullana, M. A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Meir Drexler, S., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., … Merz, C. J. (2017). Don't fear 'fear conditioning': Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. *Neuroscience & Biobehavioral Reviews*, *77*, 247–285. https://doi.org/10.1016/j.neubiorev.2017.02.026

Løseth, G. E., Eikemo, M., Isager, P., Holmgren, J., Laeng, B., Vindenes, V., Hjørnevik, T., & Leknes, S. (2018). Morphine reduced perceived anger from neutral and implicit emotional expressions. *Psychoneuroendocrinology*, *91*, 123–131. https://doi.org/10.1016/j.psyneuen.2018.02.035

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Lovibond, P. F., Saunders, J. C., Weidemann, G., & Mitchell, C. J. (2008). Evidence for expectancy as a mediator of avoidance and anxiety in a laboratory model of human avoidance learning. *Quarterly Journal of Experimental Psychology*, *61*(8), 1199–1216. https://doi.org/10.1080/17470210701503229

Luijten, M., Gillan, C. M., De Wit, S., Franken, I. H. A., Robbins, T. W., & Ersche, K. D. (2020). Goal-Directed and Habitual Control in Smokers. *Nicotine & Tobacco Research*, *22*(2), 188–195. https://doi.org/10.1093/ntr/ntz001

MacLean, R. R., Sofuoglu, M., Brede, E., Robinson, C., & Waters, A. J. (2018). Attentional bias in opioid users: A systematic review and meta-analysis. *Drug and Alcohol Dependence*, *191*, 270–278. https://doi.org/10.1016/j.drugalcdep.2018.07.012

Mahan, A. L., & Ressler, K. J. (2012). Fear conditioning, synaptic plasticity and the amygdala: Implications for posttraumatic stress disorder. *Trends in Neurosciences*, *35*(1), 24–35. https://doi.org/10.1016/j.tins.2011.06.007

Maia, T. V. (2010). Two-factor theory, the actor—Critic model, and conditioned avoidance. *Learning & Behavior*, *38*(1), 50–67. https://doi.org/10.3758/LB.38.1.50

Malafoglia, V., Ilari, S., Vitiello, L., Tenti, M., Balzani, E., Muscoli, C., Raffaeli, W., & Bonci, A. (2022). The Interplay between Chronic Pain, Opioids, and the Immune System. *The Neuroscientist*, *28*(6), 613–627. https://doi.org/10.1177/10738584211030493

Marissen, M. A. E., Franken, I. H. A., Blanken, P., Van Den Brink, W., & Hendriks, V. M. (2007). Cue Exposure Therapy for the Treatment of Opiate Addiction: Results of a Randomized Controlled Clinical Trial. *Psychotherapy and Psychosomatics*, *76*(2), 97–105. https://doi.org/10.1159/000097968

Marissen, M. A. E., Franken, I. H. A., Waters, A. J., Blanken, P., Van Den Brink, W., & Hendriks, V. M. (2006). Attentional bias predicts heroin relapse following treatment. *Addiction*, *101*(9), 1306–1312. https://doi.org/10.1111/j.1360-0443.2006.01498.x

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *32*(3), 454–465. https://doi.org/10.1016/j.neubiorev.2007.08.003

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

Martin, L., Clair, J., Davis, P., O'Ryan, D., Hoshi, R., & Curran, H. V. (2006). Enhanced recognition of facial expressions of disgust in opiate users receiving maintenance treatment. *Addiction*, *101*(11), 1598–1605. https://doi.org/10.1111/j.1360-0443.2006.01574.x

Martins, S. S., Sarvet, A., Santaella-Tenorio, J., Saha, T., Grant, B. F., & Hasin, D. S. (2017). Changes in US Lifetime Heroin Use and Heroin Use Disorder: Prevalence From the 2001-2002 to 2012-2013 National Epidemiologic Survey on Alcohol and Related Conditions. *JAMA Psychiatry*, *74*(5), 445. https://doi.org/10.1001/jamapsychiatry.2017.0113

Massaccesi, C., Willeit, M., Quednow, B. B., Nater, U. M., Lamm, C., Müller, D., & Silani, G. (2022). Opioid-blunted cortisol response to stress is associated with increased negative mood and wanting of social reward. *Neuropsychopharmacology*, *47*(10), 1798–1807. https://doi.org/10.1038/s41386-022-01283-8

Matias, S., Lottem, E., Dugué, G. P., & Mainen, Z. F. (2017). Activity patterns of serotonin neurons underlying cognitive flexibility. *eLife*, *6*, e20552. https://doi.org/10.7554/eLife.20552

McDonald, S., Darke, S., Kaye, S., & Torok, M. (2013). Deficits in social perception in opioid maintenance patients, abstinent opioid users and non-opioid users. *Addiction*, *108*(3), 566–574. https://doi.org/10.1111/add.12040

McHugh, R. K., Chase, A., Trinh, C. D., & Weiss, R. D. (2022). Are Pain and Negative Affect Coping Distinct Motives for Opioid Misuse? *Substance Use & Misuse*, *57*(6), 848–852. https://doi.org/10.1080/10826084.2022.2046094

McHugh, R. K., Fulciniti, F., Mashhoon, Y., & Weiss, R. D. (2016). Cue-induced craving to paraphernalia and drug images in opioid dependence: Opioid Cue Craving. *The American Journal on Addictions*, *25*(2), 105–109. https://doi.org/10.1111/ajad.12344

McHugh, R. K., Park, S., & Weiss, R. D. (2014). Cue-induced craving in dependence upon prescription opioids and heroin. *The American Journal on Addictions*, *23*(5), 453–458. https://doi.org/10.1111/j.1521-0391.2014.12129.x

McNally, G. P., Johansen, J. P., & Blair, H. T. (2011). Placing prediction into the fear circuit. *Trends in Neurosciences*, *34*(6), 283–292. https://doi.org/10.1016/j.tins.2011.03.005

McNally, G. P., & Westbrook, R. F. (2003). Opioid Receptors Regulate the Extinction of Pavlovian Fear Conditioning. *Behavioral Neuroscience*, *117*(6), 1292–1301. https://doi.org/10.1037/0735-7044.117.6.1292

Meier, I. M., Eikemo, M., & Leknes, S. (2021). The Role of Mu-Opioids for Reward and Threat Processing in Humans: Bridging the Gap from Preclinical to Clinical Opioid Drug Studies. *Current Addiction Reports*, *8*(2), 306–318. https://doi.org/10.1007/s40429-021-00366-8

Meulders, A., Jans, A., & Vlaeyen, J. W. S. (2015). Differences in pain-related fear acquisition and generalization: An experimental study comparing patients with fibromyalgia and healthy controls. *Pain*, *156*(1), 108–122. https://doi.org/10.1016/j.pain.0000000000000016

Miles, F. J., Everitt, B. J., & Dickinson, A. (2003). Oral cocaine seeking by rats: Action or habit? *Behavioral Neuroscience*, *117*(5), 927–938. https://doi.org/10.1037/0735-7044.117.5.927

Miles, J., & Shevlin, M. (2014). *Applying regression and correlation: A guide for students and researchers*. SAGE.

Mobbs, D., Adolphs, R., Fanselow, M. S., Barrett, L. F., LeDoux, J. E., Ressler, K., & Tye, K. M. (2019). Viewpoints: Approaches to defining and investigating fear. *Nature Neuroscience*, *22*(8), 1205–1216. https://doi.org/10.1038/s41593-019-0456-6

Morgan, M., & Marshall, J. (2013). Deficient fear recognition in regular cocaine users is not attributable to elevated impulsivity or conduct disorder prior to cocaine use. *Journal of Psychopharmacology*, *27*(6), 526–532. https://doi.org/10.1177/0269881113477708

Mowrer, O. H. (1951). Two-factor learning theory: Summary and comment. *Psychological Review*, *58*(5), 350–354. https://doi.org/10.1037/h0058956

Murphy, A., Lubman, D. I., McKie, S., Bijral, P. S., Peters, L. A., Faiz, Q., Holmes, S. E., Anderson, I. M., Deakin, B., & Elliott, R. (2018). Time-dependent neuronal changes associated with craving in opioid dependence: An fMRI study. *Addiction Biology*, *23*(5), 1168–1178. https://doi.org/10.1111/adb.12554

Myers, C. E., Rego, J., Haber, P., Morley, K., Beck, K. D., Hogarth, L., & Moustafa, A. A. (2017). Learning and generalization from reward and punishment in opioid addiction. *Behavioural Brain Research*, *317*, 122–131. https://doi.org/10.1016/j.bbr.2016.09.033

Nahin, R. L., Feinberg, T., Kapos, F. P., & Terman, G. W. (2023). Estimated Rates of Incident and Persistent Chronic Pain Among US Adults, 2019-2020. *JAMA Network Open*, *6*(5), e2313563. https://doi.org/10.1001/jamanetworkopen.2023.13563

Nahin, R. L., Sayer, B., Stussman, B. J., & Feinberg, T. M. (2019). Eighteen-Year Trends in the Prevalence of, and Health Care Use for, Noncancer Pain in the United States: Data from the Medical Expenditure Panel Survey. *The Journal of Pain*, *20*(7), 796–809. https://doi.org/10.1016/j.jpain.2019.01.003

Naqvi, N. H., & Bechara, A. (2009). The hidden island of addiction: The insula. *Trends in Neurosciences*, *32*(1), 56–67. https://doi.org/10.1016/j.tins.2008.09.009

Nees, F., Ruttorf, M., Fuchs, X., Rance, M., & Beyer, N. (2020). Brain-behaviour correlates of habitual motivation in chronic back pain. *Scientific Reports*, *10*(1), 11090. https://doi.org/10.1038/s41598-020-67386-8

Nielsen, S., Degenhardt, L., Hoban, B., & Gisev, N. (2016). A synthesis of oral morphine equivalents (OME) for opioid utilisation studies. *Pharmacoepidemiology and Drug Safety*, *25*(6), 733–737. https://doi.org/10.1002/pds.3945

Nord, C. L., Lawson, R. P., Huys, Q. J. M., Pilling, S., & Roiser, J. P. (2018). Depression is associated with enhanced aversive Pavlovian control over instrumental behaviour. *Scientific Reports*, *8*(1), 12582. https://doi.org/10.1038/s41598-018-30828-5

Nummenmaa, L., & Tuominen, L. (2018). Opioid system and human emotions. *British Journal of Pharmacology*, *175*(14), 2737–2749. https://doi.org/10.1111/bph.13812

Nury, E., Schmucker, C., Nagavci, B., Motschall, E., Nitschke, K., Schulte, E., Wegwarth, O., & Meerpohl, J. J. (2022). Efficacy and safety of strong opioids for chronic noncancer pain and chronic low back pain: A systematic review and meta-analyses. *Pain*, *163*(4), 610–636. https://doi.org/10.1097/j.pain.0000000000002423

Nutt, D. J., Lingford-Hughes, A., Erritzoe, D., & Stokes, P. R. A. (2015). The dopamine theory of addiction: 40 years of highs and lows. *Nature Reviews Neuroscience*, *16*(5), 305–312. https://doi.org/10.1038/nrn3939

O'Brien, C. P., Testa, T., O'Brien, T. J., Brady, J. P., & Wells, B. (1977). Conditioned Narcotic Withdrawal in Humans. *Science*, *195*(4282), 1000–1002. https://doi.org/10.1126/science.841320

Olney, J. J., Warlow, S. M., Naffziger, E. E., & Berridge, K. C. (2018). Current perspectives on incentive salience and applications to clinical disorders. *Current Opinion in Behavioral Sciences*, *22*, 59–69. https://doi.org/10.1016/j.cobeha.2018.01.007

Oyefeso, A., Brown, S., Chiang, Y., & Clancy, C. (2008). Self-injurious behaviour, traumatic life events and alexithymia among treatment-seeking opiate addicts: Prevalence, pattern and correlates. *Drug and Alcohol Dependence*, *98*(3), 227–234. https://doi.org/10.1016/j.drugalcdep.2008.06.004

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., & Pessiglione, M. (2012). Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. *Neuron*, *76*(5), 998–1009. https://doi.org/10.1016/j.neuron.2012.10.017

Pantazis, C. B., Gonzalez, L. A., Tunstall, B. J., Carmack, S. A., Koob, G. F., & Vendruscolo, L. F. (2021). Cues conditioned to withdrawal and negative reinforcement: Neglected but key motivational elements driving opioid addiction. *Science Advances*, *7*(15), eabf0364. https://doi.org/10.1126/sciadv.abf0364

Patterson, T. K., & Knowlton, B. J. (2018). Subregional specificity in human striatal habit learning: A meta-analytic review of the fMRI literature. *Current Opinion in Behavioral Sciences*, *20*, 75–82. https://doi.org/10.1016/j.cobeha.2017.10.005

Peciña, M., Karp, J. F., Mathew, S., Todtenkopf, M. S., Ehrich, E. W., & Zubieta, J.-K. (2019). Endogenous opioid system dysregulation in depression: Implications for new therapeutic approaches. *Molecular Psychiatry*, *24*(4), 576–587. https://doi.org/10.1038/s41380-018-0117-2

Peciña, S., & Berridge, K. C. (2013). Dopamine or opioid stimulation of nucleus accumbens similarly amplify cue-triggered 'wanting' for reward: Entire core and medial shell mapped as substrates for PIT enhancement. *European Journal of Neuroscience*, *37*(9), 1529–1540. https://doi.org/10.1111/ejn.12174

Peng, J., Sarkar, S., & Chang, S. L. (2012). Opioid receptor expression in human brain and peripheral tissues using absolute quantitative real-time RT-PCR. *Drug and Alcohol Dependence*, *124*(3), 223–228. https://doi.org/10.1016/j.drugalcdep.2012.01.013

Pennington, Z. T., Trott, J. M., Rajbhandari, A. K., Li, K., Walwyn, W. M., Evans, C. J., & Fanselow, M. S. (2020). Chronic opioid pretreatment potentiates the sensitization of fear learning by trauma. *Neuropsychopharmacology*, *45*(3), 482–490. https://doi.org/10.1038/s41386-019-0559-5

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. https://doi.org/10.1038/nature05051

Petrovic, P., Pleger, B., Seymour, B., Klöppel, S., De Martino, B., Critchley, H., & Dolan, R. J. (2008). Blocking Central Opiate Function Modulates Hedonic Impact and Anterior Cingulate Response to Rewards and Losses. *The Journal of Neuroscience*, *28*(42), 10509–10516. https://doi.org/10.1523/JNEUROSCI.2807-08.2008

Pfeiffer, A., Pasi, A., Mehraein, P., & Herz, A. (1982). Opiate receptor binding sites in human brain. *Brain Research*, *248*(1), 87–96. https://doi.org/10.1016/0006-8993(82)91150-7

Pike, A. C., & Robinson, O. J. (2022). Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, *79*(4), 313. https://doi.org/10.1001/jamapsychiatry.2022.0051

Pittig, A., Hengen, K., Bublatzky, F., & Alpers, G. W. (2018). Social and monetary incentives counteract fear-driven avoidance: Evidence from approach-avoidance decisions. *Journal of Behavior Therapy and Experimental Psychiatry*, *60*, 69–77. https://doi.org/10.1016/j.jbtep.2018.04.002

Pittig, A., Treanor, M., LeBeau, R. T., & Craske, M. G. (2018). The role of associative fear and avoidance learning in anxiety disorders: Gaps and directions for future research. *Neuroscience & Biobehavioral Reviews*, *88*, 117–140. https://doi.org/10.1016/j.neubiorev.2018.03.015

Preston, K. L., Kowalczyk, W. J., Phillips, K. A., Jobes, M. L., Vahabzadeh, M., Lin, J.-L., Mezghanni, M., & Epstein, D. H. (2018). Exacerbated Craving in the Presence of Stress and Drug Cues in Drug-Dependent Patients. *Neuropsychopharmacology*, *43*(4), 859–867. https://doi.org/10.1038/npp.2017.275

Rached, G., Vigotsky, A. D., Branco, P., Jabakhanji, R., Schnitzer, T. J., Vania Apkarian, A., & Baliki, M. N. (2022). *Long-Term Opioid Use and Emotional, Cognitive, and Sensorimotor Function in Chronic Back Pain*. Pain Medicine. https://doi.org/10.1101/2022.09.14.22279907

Reinecke, H., Weber, C., Lange, K., Simon, M., Stein, C., & Sorgatz, H. (2015). Analgesic efficacy of opioids in chronic pain: Recent meta-analyses. *British Journal of Pharmacology*, *172*(2), 324–333. https://doi.org/10.1111/bph.12634

Robbins, T. W. (2019). Commentary on Bechara et al.'s "A Neurobehavioral Approach to Addiction: Implications for the Opioid Epidemic and the Psychology of Addiction". *Psychological Science in the Public Interest*, *20*(2), 91–95. https://doi.org/10.1177/1529100619862034

Robbins, T. W., & Costa, R. M. (2017). Habits. *Current Biology*, *27*(22), R1200–R1206. https://doi.org/10.1016/j.cub.2017.09.060

Roberts, C., Apergis-Schoute, A. M., Bruhl, A., Nowak, M., Baldwin, D. S., Sahakian, B. J., & Robbins, T. W. (2022). Threat reversal learning and avoidance habits in generalised anxiety disorder. *Translational Psychiatry*, *12*(1), 216. https://doi.org/10.1038/s41398-022-01981-3

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

Ruge, J., Ehlers, M. R., Kastrinogiannis, A., Klingelhöfer-Jens, M., Koppold, A., Abend, R., & Lonsdorf, T. B. (2024). How adverse childhood experiences get under the skin: A systematic review, integration and methodological discussion on threat and reward learning mechanisms. *eLife*, *13*, e92700. https://doi.org/10.7554/eLife.92700

Sacerdote, P., Franchi, S., Gerra, G., Leccese, V., Panerai, A. E., & Somaini, L. (2008). Buprenorphine and methadone maintenance treatment of heroin addicts preserves immune function. *Brain, Behavior, and Immunity*, *22*(4), 606–613. https://doi.org/10.1016/j.bbi.2007.12.013

Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Schad, D. J., Garbusow, M., Friedel, E., Sommer, C., Sebold, M., Hägele, C., Bernhardt, N., Nebe, S., Kuitunen-Paul, S., Liu, S., Eichmann, U., Beck, A., Wittchen, H.-U., Walter, H., Sterzer, P., Zimmermann, U. S., Smolka, M. N., Schlagenhauf, F., Huys, Q. J. M., … Rapp, M. A. (2019). Neural correlates of instrumental responding in the context of alcohol-related cues index disorder severity and relapse risk. *European Archives of Psychiatry and Clinical Neuroscience*, *269*(3), 295–308. https://doi.org/10.1007/s00406-017-0860-4

Schiller, D., Levy, I., Niv, Y., LeDoux, J. E., & Phelps, E. A. (2008). From Fear to Safety and Back: Reversal of Fear in the Human Brain. *The Journal of Neuroscience*, *28*(45), 11517–11525. https://doi.org/10.1523/JNEUROSCI.2265-08.2008

Schlitt, F., Schmidt, K., Merz, C. J., Wolf, O. T., Kleine-Borgmann, J., Elsenbruch, S., Wiech, K., Forkmann, K., & Bingel, U. (2022). Impaired pain-related threat and safety learning in patients with chronic back pain. *Pain*, *163*(8), 1560–1570. https://doi.org/10.1097/j.pain.0000000000002544

Schmidt, A., Borgwardt, S., Gerber, H., Wiesbeck, G. A., Schmid, O., Riecher-Rössler, A., Smieskova, R., Lang, U. E., & Walter, M. (2014). Acute Effects of Heroin on Negative Emotional Processing: Relation of Amygdala Activity and Stress-Related Responses. *Biological Psychiatry*, *76*(4), 289–296. https://doi.org/10.1016/j.biopsych.2013.10.019

Schmidt, A., Vogel, M., Baumgartner, S., Wiesbeck, G. A., Lang, U., Borgwardt, S., & Walter, M. (2021). Brain volume changes after long-term injectable opioid treatment: A longitudinal voxel-based morphometry study. *Addiction Biology*, *26*(4), e12970. https://doi.org/10.1111/adb.12970

Schwabe, L., Tegenthoff, M., Hoffken, O., & Wolf, O. T. (2012). Simultaneous Glucocorticoid and Noradrenergic Activity Disrupts the Neural Basis of Goal-Directed Action in the Human Brain. *Journal of Neuroscience*, *32*(30), 10146–10155. https://doi.org/10.1523/JNEUROSCI.1304-12.2012

Segerdahl, A. R., Mezue, M., Okell, T. W., Farrar, J. T., & Tracey, I. (2015). The dorsal posterior insula subserves a fundamental role in human pain. *Nature Neuroscience*, *18*(4), 499–500. https://doi.org/10.1038/nn.3969

Seifert, C. L., Magon, S., Sprenger, T., Lang, U. E., Huber, C. G., Denier, N., Vogel, M., Schmidt, A., Radue, E.-W., Borgwardt, S., & Walter, M. (2015). Reduced volume of the nucleus accumbens in heroin addiction. *European Archives of Psychiatry and Clinical Neuroscience*, *265*(8), 637–645. https://doi.org/10.1007/s00406-014-0564-y

Sekutowicz, M., Guggenmos, M., Kuitunen-Paul, S., Garbusow, M., Sebold, M., Pelz, P., Priller, J., Wittchen, H.-U., Smolka, M. N., Zimmermann, U. S., Heinz, A., Sterzer, P., & Schmack, K. (2019). Neural Response Patterns During Pavlovian-to-Instrumental Transfer Predict Alcohol Relapse and Young Adult Drinking. *Biological Psychiatry*, *86*(11), 857–863. https://doi.org/10.1016/j.biopsych.2019.06.028

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential Encoding of Losses and Gains in the Human Striatum. *The Journal of Neuroscience*, *27*(18), 4826–4831. https://doi.org/10.1523/JNEUROSCI.0400-07.2007

Sifneos, P. E. (1973). The Prevalence of 'Alexithymic' Characteristics in Psychosomatic Patients. *Psychotherapy and Psychosomatics*, *22*(2–6), 255–262. https://doi.org/10.1159/000286529

Simons, L. E., Elman, I., & Borsook, D. (2014). Psychological processing in chronic pain: A neural systems approach. *Neuroscience & Biobehavioral Reviews*, *39*, 61–78. https://doi.org/10.1016/j.neubiorev.2013.12.006

Sjoerds, Z., De Wit, S., Van Den Brink, W., Robbins, T. W., Beekman, A. T. F., Penninx, B. W. J. H., & Veltman, D. J. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Translational Psychiatry*, *3*(12), e337–e337. https://doi.org/10.1038/tp.2013.107

Smoski, M. J., Salsman, N., Wang, L., Smith, V., Lynch, T. R., Dager, S. R., LaBar, K. S., & Linehan, M. M. (2011). Functional imaging of emotion reactivity in opiate-dependent borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment*, *2*(3), 230–241. https://doi.org/10.1037/a0022228

Sommer, C., Birkenstock, J., Garbusow, M., Obst, E., Schad, D. J., Bernhardt, N., Huys, Q. M., Wurst, F. M., Weinmann, W., Heinz, A., Smolka, M. N., & Zimmermann, U. S. (2020a). Dysfunctional approach behavior triggered by alcohol-unrelated Pavlovian cues predicts long-term relapse in alcohol dependence. *Addiction Biology*, *25*(1), e12703. https://doi.org/10.1111/adb.12703

Sommer, C., Birkenstock, J., Garbusow, M., Obst, E., Schad, D. J., Bernhardt, N., Huys, Q. M., Wurst, F. M., Weinmann, W., Heinz, A., Smolka, M. N., & Zimmermann, U. S. (2020b). Dysfunctional approach behavior triggered by alcohol-unrelated Pavlovian cues predicts long-term relapse in alcohol dependence. *Addiction Biology*, *25*(1), e12703. https://doi.org/10.1111/adb.12703

Soyka, M., Strehle, J., Rehm, J., Bühringer, G., & Wittchen, H.-U. (2017). Six-Year Outcome of Opioid Maintenance Treatment in Heroin-Dependent Patients: Results from a Naturalistic Study in a Nationally Representative Sample. *European Addiction Research*, *23*(2), 97–105. https://doi.org/10.1159/000468518

Spoormaker, V. I., Andrade, K. C., Schröter, M. S., Sturm, A., Goya-Maldonado, R., Sämann, P. G., & Czisch, M. (2011). The neural correlates of negative prediction error signaling in human fear conditioning. *NeuroImage*, *54*(3), 2250–2256. https://doi.org/10.1016/j.neuroimage.2010.09.042

Spunt, R. P., & Adolphs, R. (2019). The neuroscience of understanding the emotions of others. *Neuroscience Letters*, *693*, 44–48. https://doi.org/10.1016/j.neulet.2017.06.018

Srinivasan, R., Golomb, J. D., & Martinez, A. M. (2016). A Neural Basis of Facial Action Recognition in Humans. *The Journal of Neuroscience*, *36*(16), 4434–4442. https://doi.org/10.1523/JNEUROSCI.1704-15.2016

Srivastava, A. B., Mariani, J. J., & Levin, F. R. (2020). New directions in the treatment of opioid withdrawal. *The Lancet*, *395*(10241), 1938–1948. https://doi.org/10.1016/S0140-6736(20)30852-7

Strang, J., Groshkova, T., Uchtenhagen, A., Van Den Brink, W., Haasen, C., Schechter, M. T., Lintzeris, N., Bell, J., Pirona, A., Oviedo-Joekes, E., Simon, R., & Metrebian, N. (2015). Heroin on trial: Systematic review and meta-analysis of randomised trials of diamorphine-prescribing as treatment for refractory heroin addiction. *British Journal of Psychiatry*, *207*(1), 5–14. https://doi.org/10.1192/bjp.bp.114.149195

Strang, J., Volkow, N. D., Degenhardt, L., Hickman, M., Johnson, K., Koob, G. F., Marshall, B. D. L., Tyndall, M., & Walsh, S. L. (2020). Opioid use disorder. *Nature Reviews Disease Primers*, *6*(1), 3. https://doi.org/10.1038/s41572-019-0137-5

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human Pavlovian–Instrumental Transfer. *The Journal of Neuroscience*, *28*(2), 360–368. https://doi.org/10.1523/JNEUROSCI.4028-07.2008

Tanaka, S. C., Balleine, B. W., & O'Doherty, J. P. (2008). Calculating Consequences: Brain Systems That Encode the Causal Effects of Actions. *Journal of Neuroscience*, *28*(26), 6750–6755. https://doi.org/10.1523/JNEUROSCI.1808-08.2008

Terrett, G., Mercuri, K., Pizarro-Campagna, E., Hugrass, L., Curran, H. V., Henry, J. D., & Rendell, P. G. (2020). Social cognition impairments in long-term opiate users in treatment. *Journal of Psychopharmacology*, *34*(2), 254–263. https://doi.org/10.1177/0269881119875981

Tessner, K. D., Walker, E. F., Hochman, K., & Hamann, S. (2006). Cortisol responses of healthy volunteers undergoing magnetic resonance imaging. *Human Brain Mapping*, *27*(11), 889–895. https://doi.org/10.1002/hbm.20229

Thorndike, E. L. (1927). The Law of Effect. *The American Journal of Psychology*, *39*(1/4), 212. https://doi.org/10.2307/1415413

Tolomeo, S., Gray, S., Matthews, K., Steele, J. D., & Baldacchino, A. (2016). Multifaceted impairments in impulsivity and brain structural abnormalities in opioid dependence and abstinence. *Psychological Medicine*, *46*(13), 2841–2853. https://doi.org/10.1017/S0033291716001513

Trick, L., Kempton, M. J., Williams, S. C. R., & Duka, T. (2014). Impaired fear recognition and attentional set-shifting is associated with brain structural changes in alcoholic patients: Alcohol and gray matter volume. *Addiction Biology*, *19*(6), 1041–1054. https://doi.org/10.1111/adb.12175

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

Trøstheim, M., & Eikemo, M. (2024). Hyperalgesia in Patients With a History of Opioid Use Disorder: A Systematic Review and Meta-Analysis. *JAMA Psychiatry*, *81*(11), 1108. https://doi.org/10.1001/jamapsychiatry.2024.2176

Trubanova, A., Kim, I., Coffman, M. C., Bell, M. A., Richey, J. A., LaConte, S. M., Gracanin, D., & White, S. W. (2015). The Role of Perspective-Taking on Ability to Recognize Fear. *Current Research in Psychology*, *6*(2), 22–30. https://doi.org/10.3844/crpsp.2015.22.30

Upadhyay, J., Maleki, N., Potter, J., Elman, I., Rudrauf, D., Knudsen, J., Wallin, D., Pendse, G., McDonald, L., Griffin, M., Anderson, J., Nutile, L., Renshaw, P., Weiss, R., Becerra, L., & Borsook, D. (2010). Alterations in brain structure and functional connectivity in prescription opioid-dependent patients. *Brain*, *133*(7), 2098–2114. https://doi.org/10.1093/brain/awq138

Vafaie, N., & Kober, H. (2022). Association of Drug Cues and Craving With Drug Use and Relapse: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, *79*(7), 641. https://doi.org/10.1001/jamapsychiatry.2022.1240

Vaghi, M. M., Cardinal, R. N., Apergis-Schoute, A. M., Fineberg, N. A., Sule, A., & Robbins, T. W. (2019). Action-Outcome Knowledge Dissociates From Behavior in Obsessive-Compulsive Disorder Following Contingency Degradation. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(2), 200–209. https://doi.org/10.1016/j.bpsc.2018.09.014

Van Timmeren, T., Quail, S. L., Balleine, B. W., Geurts, D. E. M., Goudriaan, A. E., & Van Holst, R. J. (2020). Intact corticostriatal control of goal-directed action in Alcohol Use Disorder: A Pavlovian-to-instrumental transfer and outcome-devaluation study. *Scientific Reports*, *10*(1), 4949. https://doi.org/10.1038/s41598-020-61892-5

Vandaele, Y., & Janak, P. H. (2018). Defining the place of habit in substance use disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *87*, 22–32. https://doi.org/10.1016/j.pnpbp.2017.06.029

Verbruggen, F., Aron, A. R., Band, G. P., Beste, C., Bissett, P. G., Brockett, A. T., Brown, J. W., Chamberlain, S. R., Chambers, C. D., Colonius, H., Colzato, L. S., Corneil, B. D., Coxon, J. P., Dupuis, A., Eagle, D. M., Garavan, H., Greenhouse, I., Heathcote, A., Huster, R. J., … Boehler, C. N. (2019). A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *eLife*, *8*, e46323. https://doi.org/10.7554/eLife.46323

Vlaeyen, J. W. S., Crombez, G., & Linton, S. J. (2016). The fear-avoidance model of pain. *Pain*, *157*(8), 1588–1589. https://doi.org/10.1097/j.pain.0000000000000574

Volkow, N. D., Jones, E. B., Einstein, E. B., & Wargo, E. M. (2019). Prevention and Treatment of Opioid Misuse and Addiction: A Review. *JAMA Psychiatry*, *76*(2), 208. https://doi.org/10.1001/jamapsychiatry.2018.3126

Volkow, N. D., Wang, G.-J., Fischman, M. W., Foltin, R. W., Fowler, J. S., Abumrad, N. N., Vitkun, S., Logan, J., Gatley, S. J., Pappas, N., Hitzemann, R., & Shea, C. E. (1997). Relationship between subjective effects of cocaine and dopamine transporter occupancy. *Nature*, *386*(6627), 827–830. https://doi.org/10.1038/386827a0

Vollstädt-Klein, S., Loeber, S., Richter, A., Kirsch, M., Bach, P., Von Der Goltz, C., Hermann, D., Mann, K., & Kiefer, F. (2012). Validating incentive salience with functional magnetic resonance imaging: Association between mesolimbic cue reactivity and attentional bias in alcohol-dependent patients. *Addiction Biology*, *17*(4), 807–816. https://doi.org/10.1111/j.1369-1600.2011.00352.x

Voon, V., Joutsa, J., Majuri, J., Baek, K., Nord, C. L., Arponen, E., Forsback, S., & Kaasinen, V. (2020). The neurochemical substrates of habitual and goal-directed control. *Translational Psychiatry*, *10*(1), 84. https://doi.org/10.1038/s41398-020-0762-5

Vowles, K. E., McEntee, M. L., Julnes, P. S., Frohe, T., Ney, J. P., & Van Der Goes, D. N. (2015). Rates of opioid misuse, abuse, and addiction in chronic pain: A systematic review and data synthesis. *Pain*, *156*(4), 569–576. https://doi.org/10.1097/01.j.pain.0000460357.01998.f1

Wakeman, S. E. (2019). Harm Reduction Approaches for Opioid Use Disorder. In J. F. Kelly & S. E. Wakeman (Eds.), *Treating Opioid Addiction* (pp. 169–180). Springer International Publishing. https://doi.org/10.1007/978-3-030-16257-3_8

Walter, M., Degen, B., Treugut, C., Albrich, J., Oppel, M., Schulz, A., Schächinger, H., Dürsteler-MacFarland, K. M., & Wiesbeck, G. A. (2011). Affective reactivity in heroin-dependent patients with antisocial personality disorder. *Psychiatry Research*, *187*(1–2), 210–213. https://doi.org/10.1016/j.psychres.2010.10.017

Walter, M., Wiesbeck, G. A., Degen, B., Albrich, J., Oppel, M., Schulz, A., Schächinger, H., & Dürsteler-MacFarland, K. M. (2011). Heroin reduces startle and cortisol response in opioid-maintained heroin-dependent patients: Heroin reduces startle and cortisol response. *Addiction Biology*, *16*(1), 145–151. https://doi.org/10.1111/j.1369-1600.2010.00205.x

Wang, S.-C., & Maher, B. (2019). Substance Use Disorder, Intravenous Injection, and HIV Infection: A Review. *Cell Transplantation*, *28*(12), 1465–1471. https://doi.org/10.1177/0963689719878380

Wang, Z.-X., Zhang, J. X., Wu, Q.-L., Liu, N., Hu, X.-P., Chan, R. C. K., & Xiao, Z.-W. (2010). Alterations in the processing of non-drug-related affective stimuli in abstinent heroin addicts. *NeuroImage*, *49*(1), 971–976. https://doi.org/10.1016/j.neuroimage.2009.08.020

Wardle, M. C., Bershad, A. K., & De Wit, H. (2016). Naltrexone alters the processing of social and emotional stimuli in healthy adults. *Social Neuroscience*, *11*(6), 579–591. https://doi.org/10.1080/17470919.2015.1136355

Wardle, M. C., Fitzgerald, D. A., Angstadt, M., Rabinak, C. A., De Wit, H., & Phan, K. L. (2014). Effects of oxycodone on brain responses to emotional images. *Psychopharmacology*, *231*(22), 4403–4415. https://doi.org/10.1007/s00213-014-3592-4

Ware, O. D., & Dunn, K. E. (2023). Clinically meaningful individual differences in opioid withdrawal expression. *Experimental and Clinical Psychopharmacology*, *31*(6), 1005–1009. https://doi.org/10.1037/pha0000654

Warner-Smith, M., Darke, S., & Day, C. (2002). Morbidity associated with non-fatal heroin overdose. *Addiction*, *97*(8), 963–967. https://doi.org/10.1046/j.1360-0443.2002.00132.x

Wasan, A. D., Butler, S. F., Budman, S. H., Fernandez, K., Weiss, R. D., Greenfield, S. F., & Jamison, R. N. (2009). Does Report of Craving Opioid Medication Predict Aberrant Drug Behavior Among Chronic Pain Patients? *The Clinical Journal of Pain*, *25*(3), 193–198. https://doi.org/10.1097/AJP.0b013e318193a6c4

Wassum, K. M., Cely, I. C., Maidment, N. T., & Balleine, B. W. (2009). Disruption of endogenous opioid activity during instrumental learning enhances habit acquisition. *Neuroscience*, *163*(3), 770–780. https://doi.org/10.1016/j.neuroscience.2009.06.071

Weber, S. C., Beck-Schimmer, B., Kajdi, M.-E., Müller, D., Tobler, P. N., & Quednow, B. B. (2016). Dopamine D2/3- and μ-opioid receptor antagonists reduce cue-induced responding and reward impulsivity in humans. *Translational Psychiatry*, *6*(7), e850–e850. https://doi.org/10.1038/tp.2016.113

Webster, L. R. (2017). Risk Factors for Opioid-Use Disorder and Overdose. *Anesthesia & Analgesia*, *125*(5), 1741–1748. https://doi.org/10.1213/ANE.0000000000002496

Weiss, R. D., Potter, J. S., Griffin, M. L., McHugh, R. K., Haller, D., Jacobs, P., Gardin, J., Fischer, D., & Rosen, K. D. (2014). Reasons for opioid use among patients with dependence on prescription opioids: The role of chronic pain. *Journal of Substance Abuse Treatment*, *47*(2), 140–145. https://doi.org/10.1016/j.jsat.2014.03.004

White, J. M., & Irvine, R. J. (1999). Mechanisms of fatal opioid overdose. *Addiction*, *94*(7), 961–972. https://doi.org/10.1046/j.1360-0443.1999.9479612.x

Winkelman, T. N. A., Chang, V. W., & Binswanger, I. A. (2018). Health, Polysubstance Use, and Criminal Justice Involvement Among Adults With Varying Levels of Opioid Use. *JAMA Network Open*, *1*(3), e180558. https://doi.org/10.1001/jamanetworkopen.2018.0558

Wise, R. (2002). Combining fMRI with a Pharmacokinetic Model to Determine Which Brain Areas Activated by Painful Stimulation Are Specifically Modulated by Remifentanil. *NeuroImage*, *16*(4), 999–1014. https://doi.org/10.1006/nimg.2002.1146

Wise, R. A., & Koob, G. F. (2014). The Development and Maintenance of Drug Addiction. *Neuropsychopharmacology*, *39*(2), 254–262. https://doi.org/10.1038/npp.2013.261

Wollman, S. C., Alhassoon, O. M., Hall, M. G., Stern, M. J., Connors, E. J., Kimmel, C. L., Allen, K. E., Stephan, R. A., & Radua, J. (2017). Gray matter abnormalities in opioid-dependent patients: A neuroimaging meta-analysis. *The American Journal of Drug and Alcohol Abuse*, *43*(5), 505–517. https://doi.org/10.1080/00952990.2016.1245312

Wollman, S. C., Alhassoon, O. M., Stern, M. J., Hall, M. G., Rompogren, J., Kimmel, C. L., & Perez-Figueroa, A. M. (2015). White matter abnormalities in long-term heroin users: A preliminary neuroimaging meta-analysis. *The American Journal of Drug and Alcohol Abuse*, *41*(2), 133–138. https://doi.org/10.3109/00952990.2014.985829

Worbe, Y., Savulich, G., De Wit, S., Fernandez-Egea, E., & Robbins, T. W. (2015). Tryptophan Depletion Promotes Habitual over Goal-Directed Control of Appetitive Responding in Humans. *International Journal of Neuropsychopharmacology*, *18*(10), pyv013. https://doi.org/10.1093/ijnp/pyv013

World Health Organization. (2023). *WHO Model List of Essential Medicines—23rd list*. https://www.who.int/publications/i/item/WHO-MHP-HPS-EML-2023.02

Xia, Y., Gurkina, A., & Bach, D. R. (2019). Pavlovian-to-instrumental transfer after human threat conditioning. *Learning & Memory*, *26*(5), 167–175. https://doi.org/10.1101/lm.049338.119

Younger, J. W., Chu, L. F., D'Arcy, N. T., Trott, K. E., Jastrzab, L. E., & Mackey, S. C. (2011). Prescription opioid analgesics rapidly change the human brain. *Pain*, *152*(8), 1803–1810. https://doi.org/10.1016/j.pain.2011.03.028

Zhao, K., Yan, W.-J., Chen, Y.-H., Zuo, X.-N., & Fu, X. (2013). Amygdala Volume Predicts Inter-Individual Differences in Fearful Face Recognition. *PLoS ONE*, *8*(8), e74096. https://doi.org/10.1371/journal.pone.0074096

Zilverstand, A., Huang, A. S., Alia-Klein, N., & Goldstein, R. Z. (2018). Neuroimaging Impaired Response Inhibition and Salience Attribution in Human Drug Addiction: A Systematic Review. *Neuron*, *98*(5), 886–903. https://doi.org/10.1016/j.neuron.2018.03.048

Zubieta, J. (2000). Buprenorphine-Induced Changes in Mu-Opioid Receptor Availability in Male Heroin-Dependent Volunteers A Preliminary Study. *Neuropsychopharmacology*, *23*(3), 326–334. https://doi.org/10.1016/S0893-133X(00)00110-X

Zwosta, K., Ruge, H., Goschke, T., & Wolfensteller, U. (2018). Habit strength is predicted by activity dynamics in goal-directed brain systems during training. *NeuroImage*, *165*, 125–137. https://doi.org/10.1016/j.neuroimage.2017.09.062

Acknowledgements

First, I would like to thank everyone who participated in the study, many of whom have been through so much pain - physical or otherwise - and have greatly inspired me with their relentless hope of a better tomorrow for themselves and the people around them. This thesis is dedicated to them.

I would like to thank my supervisor Karen Ersche for giving me the opportunity to discover the thrill of science, for all the knowledge and practical advice she passed on to me, and for her guidance throughout the last four years.

I would like to thank Christian Büchel who gave me a home at the ISN when I first arrived in Hamburg during a COVID-struck autumn and taught me a great deal about the brain and the intricacies of researching it. I would also like to thank all members of Christian's group, past or present, for the lively discussions that have greatly expanded my scientific horizon.

This would not have been possible without all my wonderful colleagues and friends at the ISN, who have kept me company and endured my whining and my over-enthusiasm (sometimes both within a matter of seconds) throughout the last four years: Christoph, Maren, and Jan; Madeleine, Jan, and Leonie; Ole, Lars, Leo, and Marie. Thank you and a whole lot of love for all of you. I would like to single out Janne for her great support and encouragement during the writing of this thesis. I would also like to thank Carolina for providing such reliable structural support and, impressively, always keeping a cool head about her.

I would also like give all my love to my friends, holding me down and making sure I don't lose touch with reality: Vincent, Schelli, Hannes, Niklas, Jesper, Carla, Leonie, Leona. Also Amke, who has been tirelessly by my side, being more loving, patient, and gentle with me than I probably deserved, and all this while being the single coolest cat out there. Also also, my parents, Claudia and Martin, to whom I am more grateful than I could ever express.

Curriculum Vitae

Caspar Jakob Geißler

Psychologist, M.Sc.

born January 4th 1996, Flensburg, Germany

contact c.geissler@uke.de

currrently

October 2020 – currently

Research associate

Department of Systems Neuroscience
University Medical Centre Hamburg-Eppendorf

Project:
Effects of Chronic Opioid Use on Brain and Behaviour

Supervision:
Prof. Dr. Karen Ersche
Department of Psychiatry, University of Cambridge

January 2021 – currently

PhD student

Non-medical PhD-Programm
Medical faculty - Universität Hamburg

Thesis:
Aversion processing following chronic opioid use with and without addiction

Supervision :
Prof. Dr. Christian Büchel
Department of Systems Neuroscience
University Medical Centre Hamburg-Eppendorf

past (selection)

November 2018 – June 2020

Research assistant
General Psychology, Albert-Ludwigs-Universität Freiburg

September 2017 – March 2019

Intern
Centre for Interdisciplinary Addiction Research,
University Medical Centre Hamburg-Eppendorf

Curriculum Vitae (continued)

Education

October 2017 – September 2020	**Master of Science** **Clincal Psychology, neuroscience, and rehabilitation sciences** Department of Psychology Albert-Ludwigs-Universität Freiburg Master's thesis: The Influence of Facial Attractiveness on Trust in Men and Women
October 2014 – September 2017	**Bachelor of Science** **Psychology** Department of Psychology Albert-Ludwigs-Universität Freiburg Bachelor's thesis: Recognising students' dysfunctional beliefs in future teachers
September 2005 – June 2013	**Abitur** Karl-Friedrich-Gymnasium Mannheim

Publications & posters (selection)

August 2023	**Geißler, C.** & Ersche, K. (2023). *Aversive learning following regular opioid use* (Poster) Meeting of the European Behavioural Pharmacology Society, Mannheim, Deutschland.
October 2021	**Geißler, C.** & Ersche, K. (2021). *What renders people vulnerable to developing opioid addiction?* (Poster) Meeting of the European College for Neuropsychopharmacology, Lissabon, Portugal.
August 2019	**Geißler, C.** & Schredl, M. (2020). College students' erotic dreams: Analysis of content and emotional tone. *Sexologies, 29*(1), 11-17.
July 2019	Schredl, M., **Geißler, C.**, & Göritz, A. S. (2019). Factors influencing the frequency of erotic dreams: an online study. *Psychology & Sexuality*, *10*(4), 316-324.

Caspar Geißler – c.geissler@uke.de

Eidesstattliche Erklärung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe, insbesondere ohne entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten, verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe. Das gilt insbesondere auch für alle Informationen aus Internetquellen.

Soweit beim Verfassen der Dissertation KI-basierte Tools („Chatbots“) verwendet wurden, versichere ich ausdrücklich, den daraus generierten Anteil deutlich kenntlich gemacht zu haben. Die „Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG“ aus September 2023 wurde dabei beachtet.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich damit einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Hamburg, 24. Januar 2025

(Datum) (Unterschrift)