PSYCHOPATHIC TRAITS AND MECHANISMS OF ANTISOCIAL DECISION-MAKING

Josi Driessen

DONDERS SERIES

The work presented in this thesis has been supported by (i) the European Union Innovative Medicine Initiative (IMI) grant PRISM (n° 115916), (ii) the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 603016 (MATRICS), and (iii) the Department of Cognitive Neuroscience of Radboudumc.

ISBN
978-94-6421-220-4

Design/lay-out
Promotie In Zicht (www.promotie-inzicht.nl)

Print
Ipskamp Printing

PSYCHOPATHIC TRAITS AND MECHANISMS OF ANTISOCIAL DECISION-MAKING

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag 6 juli 2021
om 12.30 uur precies

door

Josi Maria Anna Driessen
geboren op 8 december 1989
te Nijmegen

Promotoren
Prof. dr. R.P.C. Kessels
Prof. dr. J. K. Buitelaar

Copromotoren
Dr. I. A. Brazil
Dr. J. C. Glennon (University College Dublin, Ierland)

Manuscriptcommissie
Prof. dr. K. Roelofs
Dr. R. B. Mars
Prof. dr. E. Viding (University College London, Verenigd Koninkrijk)

'It is not easy to convey this concept, that of a biological organism, outwardly intact, showing excellent peripheral function, but centrally deficient or disabled, in such a way that abilities, excellent at the only levels where we can formally test them, cannot be utilized consistently for sane purposes or prevented from regularly working toward self-destructive and seriously pathological results.'

Hervey Cleckley, 1955

Contents

1

General introduction

1.1 Introduction

Most people are familiar with the term ‘psychopath’, which is often used to label someone who is “morally insane”, “murderous”, and “coldhearted”. Its popularity among the general public is reflected by the wide variety of novels, movies, tv series, and computer games, in which a notorious ‘psychopath’ fulfills the starring role. Importantly, their characterizations often only partly relate to the concept of psychopathy as it is defined in psychiatry, criminology, and in research. Psychopathy is a personality disorder that has received a great deal of attention over the past centuries. In the history of psychiatry, psychopathy was first used as a generic term for any severe psychiatric illness. Only later, it became linked to what is now known as psychopathy. *The Mask of Sanity* by the American psychiatrist Hervey M. Cleckley, first published in 1941 and with revised editions appearing for several decades, is considered a seminal work that marked the beginning of the modern clinical construct of psychopathy. Cleckley was not the first to recognize the characteristics of psychopathy, as over the preceeded two centuries many other scholars, including Pinel, Prichard, Koch and Kraepelin, have theorized about psychopathy or conditions that are much alike (Arrigo & Shipley, 2001). Nevertheless, Cleckley was the first to lay out in detail the key characteristics of the disorder, and moreover, his conceptualization of psychopathy became more influential and familiar to laypersons and academics than those of other scholars (Crego & Widiger, 2015).

Cleckley characterized psychopathy based on sixteen criteria he believed formed the core features of psychopathy. These criteria are (1) superficial charm and good “intelligence”, (2) absence of delusions and other signs of irrational thinking, (3) absence of nervousness and psychoneurotic manifestations, (4) unreliability, (5) untruthfulness and insincerity, (6) lack of remorse or shame, (7) inadequately motivated antisocial behavior, (8) poor judgment and failure to learn by experience, (9) pathological egocentricity and incapacity for love, (10) general poverty in major affective reactions, (11) specific loss of insight, (12) unresponsiveness in general interpersonal relations, (13) fantastic and uninviting behavior with drink and sometimes without, (14) suicide rarely carried out, (15) sex life impersonal, trivial, and poorly integrated, and (16) failure to follow any life plan (Cleckley, 1941/1976, pp. 338–339). Increasing interest in the field has led to different conceptualizations of psychopathy, and the exact definition of the construct is stil a matter of debate (Hare & Neumann, 2010; Miller & Lynam, 2015; Skeem & Cooke, 2010). However, the idea that emotional disturbances and lack of morality play key roles in psychopathy is still prominently present.

1.2 The construct of psychopathy

Although Cleckley's criteria were often treated as if they reflect psychopathy as a unitary construct, ample evidence suggested that psychopathy consists of a constellation of traits that vary along a continuum (DeLisi, 2016). Cleckley's criteria did not allow to measure the extent to which individuals matched the criteria. To overcome this limitation, Hare developed the Psychopathy Checklist (PCL, 1980) and its successor, the Psychopathy Checklist—Revised (PCL-R; Hare, 1991/2003), which were greatly inspired by Cleckley's work. The PCL-R covers twenty items that are used to assess the presence of psychopathic traits in adults based on a semi-structured interview and an extensive review of collateral file information. Factor analyses demonstrated that the PCL-R distinguishes between two factors; disturbed interpersonal-affective behavior (Factor 1; F1) and impulsive-antisocial (Factor 2; F2) traits. Where Factor 1 is associated with personality features (e.g. lack of empathy, lack of guilt, superficial charm and pathological lying) that capture the core features that are unique to psychopathy, Factor 2 describes behaviors (e.g. parasitic lifestyle, impulsive behavior, poor behavioral controls and juvenile delinquency) that represent a more general set of antisocial tendencies that can be found across several subtypes of antisocial individuals (Hansen, Johnsen, Thornton, Waage, & Thayer, 2007; Hare, 2003; Ross & Rausch, 2001). The two-factor model proposed by Hare dominated the literature on psychopathy for a long time (Harpur, Hakstian, & Hare, 1988).

However, alternative conceptualizations have been developed in parallel to the two-factor model of the PCL-R (Cook & Michie, 2001; Feilhauer & Cima, 2013; Lilienfeld & Andrews, 1996; Patrick, Fowles, & Kreuger, 2009). For instance, Cooke and Michie (2001) argued that antisociality is a consequence of affective-emotional impairments and is, therefore, of secondary importance to the definition of psychopathy. They proposed a three-factor hierarchical model of psychopathy that emphasized criminal behavior to a lesser extent (Cooke & Michie, 1997). According to this model, the construct is underpinned by three correlated factors; 'arrogant and deceitful interpersonal style', 'deficient and affective experience', and 'impulsive and irresponsible behavioral style', and this was supported by confirmatory factor analyses on several large data sets (Cooke & Michie, 1997). Hare and colleagues disputed a three-factor model and cited the importance of differentiating between antisocial and criminal behavior (Hare & Neumann, 2008; Hare & Neumann, 2010; Neumann, Vitacco, Hare, & Wupperman, 2005). They argued that antisociality (generalized rule breaking) is an intrinsic component of the psychopathy construct. A further refinement of Hare's two-factor model distinguishes two individual facets for each factor, where F1 is divided into an 'interpersonal' and an 'affective' facet and F2 is divided into a 'lifestyle' and an 'antisocial' facet. The 'interpersonal' facet measures arrogant and deceitful interpersonal style, which is characterized by superficial charm, grandiosity, manipulative behavior and deceitfulness. The 'affective' facet measures the degree of deficient affective experience which encompasses callousness, lack of empathy, failure to accept responsibility and lack of remorse or guilt. The 'lifestyle' facet measures impulsive-irresponsible behavioral style which is characterized by impulsivity, boredom, sensation seeking, parasitic lifestyle, irresponsibility, and lack of goals. The 'antisocial' facet encompasses aggressiveness, early behavior problems, juvenile delinquency and criminal versatility (Hare & Neumann, 2005). Although the items of the PCL-R are generally interrelated, confirmatory factor analysis provided evidence for four uni-dimensional facets. Factor 1 and Factor 2 are still used and referred to as the interpersonal-affective factor and the impulsive-antisocial, respectively, in order to include both the two superordinate factors and the presence of the four facets (Figure 1.1). Hare's four-factor model received extensive support from several large PCL-based studies including a wide variety of samples (Hill, Neumann, & Rogers, 2004; Kosson, Cyterski, Steuerwald, Neumann, & Walker-Matthews, 2002; Neumann, Hare, & Newman, 2007; Olver, Neumann, Wong, & Hare, 2013; Vitacco, Neumann, & Jackson, 2005), and is the most validated and popular method for measuring psychopathic traits (Anderson & Kiehl, 2012).

Psychopathy is a particularly dangerous and malignant constellation of personality traits and an extensive amount of research has been devoted to unravel its nature and correlates. Nevertheless, psychopathy is not recognized as a disorder in the

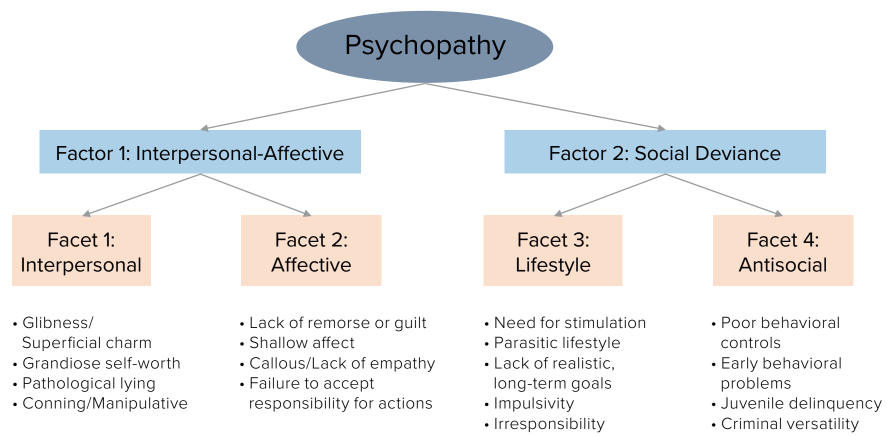


Figure 1.1 Hare's four-factor model with the two factors Interpersonal-Affective (Factor 1) and Social deviance (i.e. Lifestyle-Antisocial, Factor 2) and the four facets Interpersonal, Affective, Lifestyle, and Antisocial.

American Psychiatric Association's (APA) Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 2013). The Cleckley-oriented conceptions of psychopathy that were present in the first two editions of the DSM were replaced by Antisocial Personality Disorder (ASPD) in the later editions. ASPD was intended to be equated to psychopathy, but this has been intensly debated. One common criticism is that ASPD is almost exclusively based on behavioral characteristics, but hardly recognizes the underlying personality traits. Importantly, research has suggested that psychopathy can also be present in the absence of antisocial behavior (e.g. Hall & Benning, 2006), and in addition, psychopathic individuals are known for their ability to deceive and manipulate therapists and cover their crimes (Hare, Forth, & Hart, 1989). Thus, predominantly focusing on the antisocial aspects and criminal behavior could cause an important subgroup of individuals with psychopathic traits to remain undetected. This also suggests that individuals with psychopathic and antisocial traits can be classified into distinct subtypes, which is another topic of debate in psychopathy literature.

1.3 Subtypes of psychopathy

In line with Hare's conceptualization, multiple proposals on how to subtype individuals based on psychopathic features are based on the distinction between primary psychopathy and secondary psychopathy. Early proposals suggested that primary psychopathy is a genetically determined affective *deficit*, while secondary psychopathy reflects an affective *disturbance* caused by the influence of a broader set of factors, including early psychosocial learning. Primary psychopathy is characterized by low anxiety and goal-directed behavior, while secondary psychopathy was characterized by high anxiety, and impulsive actions (Brazil & Cima, 2016; Brazil, van Dongen, Maes, Mars, & Baskin-Sommers, 2018; Poythress, Skeem, & Lilienfeld, 2006). An influential theory that inspired the formulation of other hypotheses about subtyping of psychopathy was Gray's reinforcement sensitivity theory (Gray, 1970). This theory described the Behavioral Inhibition System (BIS) and the Behavioral Activation System (BAS) as drivers of instrumental behavior. The BIS system is involved in inhibition of goal-directed behavior in reponse to aversive stimuli, while the BAS system is responsible for initiation and modulation of behavior that is driven by factors such as reward or lack of punishment. Based on this theory, Lykken (1995) and Fowles (1980) formulated a distinction of primary psychopathy, characterized by low levels of BIS, and secondary psychopathy, associated with hyperactive BAS. Empirical support for a similar distinction was provided by a study that showed that primary psychopathy was linked to weak BIS and normal BAS combined with low trait anxiety, while

secondary psychopathy was associated with strong BAS scores combined with high trait anxiety, while the role of BIS in this subtype remained unclear (Newman, MacCoon, Vaughn, & Sadeh, 2005). Another study supported this distinction and found support for an association between interpersonal-affective traits (Factor 1 of the PCL-R) and the BIS after controlling for the presence of impulsive-antisocial traits (Factor 2 of the PCL-R), and between Factor 2 and BAS after controlling for Factor 1 traits (Wallace, Malterer, & Newman, 2009). This idea was followed up by other empirical studies that proposed similar subtypes of psychopathy and further characterized these subtypes based on related personality and behavioral characteristics (e.g., Driessen et al., 2018 (chapter 2); Patrick, Hicks, Krueger, & Lang, 2005; Skeem, Poythress, Edens, Lilienfeld, & Cale, 2003).

1.4 Psychopathic traits in the general population

The majority of published literature on psychopathy has focused on psychopathy in criminal offenders (Thompson, Ramos, & Willett, 2014). Interestingly, the early conceptualizations of psychopathy were not at all focused on the criminal aspects of the disorder. In one of his seminal reports, Cleckley (1941) recognized that many psychopaths never became involved with the criminal justice system. Some researchers suggested that Hare's model underexposed the adaptiveness of certain psychopathic traits which could account for a successful adjustment to society (Patrick & Bernat, 2009). These adaptive traits were also recognized in one of Cleckley's early reports, in which he highlighted different expressions of a similar core deficit of psychopathy. On one hand, the callous-unemotional traits were described as a prerequisite for disinhibition, and antisocial and destructive behavior. However, on the other hand, some individuals scoring high on these behaviors appeared to be equipped with advanced social skills and use them in their advantage (e.g. to manipulate). Nowadays, the existence of psychopathic personality traits in the general population has been widely recognized (Gao & Raine, 2010; Levenson, Kiehl, & Fitzpatrick, 1995), and individuals are indicated as psychopathic based on the PCL-R are thought to express these traits to an extreme extent (Hare & Neumann, 2008).

Since scoring criteria for several items of the PCL-R strongly refer to criminal behavior, its value for measuring psychopathic traits in the general population has been criticized. Nevertheless, there are several methods for assessment of psychopathic traits in non-offender samples, such as the Self-Report Psychopathy scale (SRP; Paulhus, Neumann, & Hare, 2015), the Triarchic Psychopathy Measure (TriPM; Paulhus et al., 2015), the Levenson's Self-Report Psychopathy Scale (LSRP; Levenson et al., 1995), and the Psychopathic Personality Inventory (PPI; Lilienfeld

& Widows, 2005). Each method has its particular strengths and weaknesses (Lilienfeld & Fowler, 2006). Given the extensive amount of support for the four-factor model with offender samples (Neumann et al., 2007) and psychiatric samples (Jackson, Neumann, & Vitacco, 2007), but also in a large community sample (Hare & Neumann, 2008), the theoretical proximity of the SRP and the PCL-R offers an advantage over all other self-report inventories in psychopathy research.

An increasing interest of assessing psychopathy in the general population also led to the investigation as to whether the same factor structure that has been found in criminal populations may apply to community samples (Williams, Paulhus, & Hare, 2007). A few studies examined the distribution of psychopathic traits in community samples and distinguished profiles that were also described in the offender literature (Falkenbach, Poythress, Falki, & Manchak, 2007; Hare & Neumann, 2008; Lee, Salekin, & Iselin, 2010). A review on psychopathic traits in non-offender populations proposed a neurobiological model that distinguished 'successful' and 'unsuccessful' psychopaths (Gao & Raine, 2010). According to this model, impairments in brain structure and function are thought to underlie cognitive and affective dysfunction and increased violent offending in unsuccesful psychopaths. As a consequence, they are more likely to end up in penitentiary institutions. On the contrary, an intact neurobiological functioning combined with strong PCL-R interpersonal and affective features of successful (or adaptive or non-criminal) psychopaths are suggested to be advantageous in the business and corporate world (Babiak, Hare, & McLaren, 2006; Hall & Benning, 2006; Lykken, 1995). Nevertheless, this 'success' is often temporary, dependent on the circumstances, and has a negative impact on others (de Oliveira-Souza et al., 2008). Thus, although some psychopathic traits can be advantageous on a personal level, people that have psychopathic tendencies and are prone to express antisocial behavior, are a potential threat to society. Moreover, their decisions can cause great suffering to their families, friends, and colleagues (Mathieu & Babiak, 2016). Therefore, it is important to learn more about the presence of psychopathic traits in the general community (Anderson & Kiehl, 2012).

1.5 Interim summary

Psychopathy is a concept that has been studied for a long time. In the past century, psychiatrists started to lay out in detail the key characteristics of the disorder. Reseachers built on this knowledge and further structured the concept and developed methods to assess the presence of psychopathic traits. Over the years, a variety of different conceptualizations of psychopathy have been proposed. Nowadays, the PCL(-R) is the most commonly measure of psychopathy. Although

the PCL-R was designed to measure psychopathic traits in clinical and offender populations, its factor structure was also found to be valid in community samples. Self-report questionnaires have been developed to ease testing of large community samples, and their use has been validated in a variety of samples. The existence of multiple conceptualizations highlights the lack of consensus about the construct of psychopathy. Many researchers recognize that a personality-based characterization of psychopathy is insufficient and argue that insight into the neural and cognitive mechanisms underlying their behavior is necessary to further understand the construct and to be able to distinguish subtypes. Knowledge about these mechanisms is also important for the development of better and personalized treatment interventions, as currently their effectiveness have been disappointing (Baskin-Sommers, Curtin, & Newman, 2015; Brazil et al., 2018).

1.6 A neurocognitive perspective on psychopathy

Recent technological advances have contributed to a remarkable progress into our understanding of cognitive and neurobiological mechanisms that are driving abnormal behavior in various personality disorders, including psychopathy. This has led to an increase in perspectives that aim to explain psychopathy not only based on its personality and behavioral characteristics, but also on the underlying cognitive and neurobiological components. Several of these accounts couched their explanations primarily at either the cognitive level (e.g. response modulation theory; Newman, 1998; Patterson & Newman, 1993) or at the neural level (e.g. the frontal lobe dysfunction hypothesis; Elliott, 1978; Gorenstein, 1982; Moffitt, 1993; Raine, 2002). However, these single level approaches could not provide a complete account of the range of impairments that characterized individuals with high levels of psychopathic traits. Conversely, the Integrated Emotions System model (IES; Blair, 2005, 2013) integrates the cognitive and neural levels and provides an explanation for how specific functional impairments in a restricted set of interacting neural systems give rise to psychopathy-related behavior (Figure 1.2). According to this model, psychopathic traits can be characterized by core impairments in emotional empathy, particularly in the processing of distress cues, and core impairments in decision-making, more specifically in reinforcement learning and outcome representation. These affective and cognitive impairments have been found to be associated with dysfunction in the amygdala, the ventromedial prefrontal cortex (vmPFC), and the striatum. Although other neural regions, including the dorsomedial prefrontal cortex and the anterior insular cortex, were proposed to be implicated as well, the amygdala, vmPFC, and the striatum regions play a central role in this theory of psychopathy.

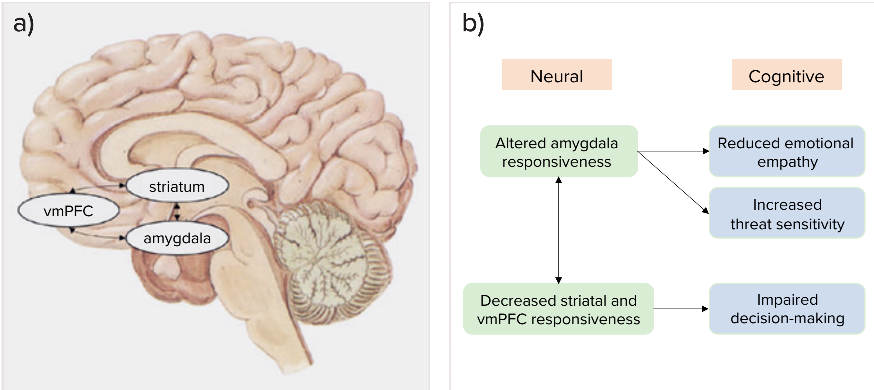


Figure 1.2 a) Overiew of the core neural regions implicated in psychopathic traits; the amygdala, the striatum, and the ventromedial prefrontal cortex (vmPFC). **b)** Model representation of the interactions between the neural regions and the associated cognitive impairments. Adapted from (Blair, 2013).

The amygdala is a subcortical structure that is involved in the acquisition of stimulus-reinforcement associations and the recognition of emotionally salient information, including threat detection (Davis & Whalen, 2001). A division between the basolateral and the central nuclei of the amygdala was already proposed almost a century ago (Johnston, 1923), and this distinction is still widely implemented in modern research. In line with expectations based on behavioral and psycho-physiological findings, a large amount of evidence has indicated the presence of structural (Boccardi et al., 2011; Ermer, Cope, Nyalakanti, Calhoun, & Kiehl, 2012; Kaya, Yildirim, & Atmaca, 2020; Yang, Raine, Colletti, Toga, & Narr, 2010) and functional disturbances in the amygdala of individuals with high levels of psycho-pathic traits (Dolan & Fullam, 2009; Harenski, Harenski, Shane, & Kiehl, 2010; Harenski, Kim, & Hamann, 2009; Johanson, Vaurio, Tiihonen, & Lähteenvuo, 2020, Kiehl et al., 2001). The structural differences typically indicate grey-matter volume reductions (Ermer et al., 2012; Yang et al., 2010), and the functional disturbances often concern diminished activation during aversive conditioning (Birbaumer et al., 2005), facial emotion recognition (Dolan & Fullam, 2009; Gordon, Baird, & End, 2004), and moral judgement (Harenski, Harenski, & Kiehl, 2014; Marsh & Cardinale, 2014). An important aspect of the amygdala in relation to psychopathy is its involvement in establishing stimulus-reinforcement associations, in particular in learning about cues that signal distress in others (e.g. fearful, anger or sad facial expressions) (Dawel, O'Kearney, McKone, & Palermo, 2012; Marsh & Blair, 2008; Vieira, Tavares, Marsch, & Mitchell, 2017). Individuals that are unaware of these

relationships are more likely to commit an action that causes harm to others and are less likely to inhibit harmful behavior when confronted with distress in others. The exact neurobiological underpinnings of such behaviors remain unclear, but abnormalities in endogenous testosterone levels were proposed to be an underlying source of impaired functioning in the amygdala, because of its association with approach-related behaviors including reward-seeking (Baskin-Sommers, Wallace, MacCoon, Curtin, & Newman, 2010; Buckholtz et al., 2010), dominance (Archer & Webb, 2006), and aggression (Kiehl & Hoffman, 2011). Given that these behaviors are characteristic to psychopathy as well, testosterone is one of the factors that has been taken into account in psychopathy literature (Yilderim & Derksen, 2012).

Like the amygdala, the prefrontal cortex has featured prominently in theories of psychopathy, because of its association with poor moral judgement and impulsivity that also characterizes psychopathic decision-making (Blair, 2007; Kim & Lee, 2011). The vmPFC remains the most common prefrontal region implicated in recent neuroimaging investigations of psychopathy. The structural differences related to psychopathy typically indicate reduced grey-matter and cortical thickness (Boccardi et al., 2011; Tiihonen et al., 2008; Yang et al., 2010), and functional disturbances reflect reduced activation during social cooperation (Rilling et al., 2007), emotion integration (Müller et al., 2008), and moral reasoning (Pujol et al., 2012). In addition, atypical functional connectivity with the amygdala has been suggested (Waller et al., 2019; Marsh et al., 2011), which indicates reduced interactions between these two regions.

Lastly, a large number of brain imaging studies have identified differences in the structure and functioning of the striatum in individuals with high levels of psychopathic traits and antisociality (Glenn & Yang, 2012; Korponay et al., 2017). The striatum is a subcortical region that is involved in the processing of reward-related information and has been linked to reward seeking and impusivity, which are both prevalent in psychopathy. In addition, it has been found to play a role in stimulus-response and response-outcome learning (Horvitz, 2009). Neuroimaging studies found an increased striatal volume (Glenn, Raine, Yaralian, & Yang, 2010), and increased functioning in individuals with psychopathy. In particular, the striatum was found to be hyperactive in a situation where a rewarding stimulus changes in a non-rewarding stimulus, resulting in continuous responding to a stimulus that is no longer rewarding. This could reflect a decreased ability to respond to changes in the environment, and may be involved in impulsive and antisocial behavior (Glenn & Yang, 2012).

According to Blair's IES model, structural and functional disturbances in these three key areas heavily contribute to the core impairments in empathy and decision-making that are associated with psychopathy. Empathic reactions can be elicited in response to cues that signal emotions others, such as facial or vocal expressions, body posture, but also images or text. Empathy has an important communicatory function and contributes to prosocial behavior and socialization (Blair, 2003; Weinstein, Feldman, Goodman, & Markowitz, 1972). Essential components of empathy have been found to be impaired in individuals with psychopathic traits. Meta-analytic findings demonstrated that psychopathy was associated with altered emotion recognition (Dawel et al., 2012) and reduced responsiveness (Brook, Brieman, & Kosson, 2013; Jensen et al., 2001; Lykken, 1957; Patrick, 2018) across different modalities. These recognition and processing deficits were predominantly found in distress cues, such as fear and sadness (Marsh & Blair, 2008), and are associated with amygdala dysfunction. Importantly, individuals with elevated psychopathic tendencies tend to engage in violent behavior that causes great harm and pain to others (Mathieu & Babiak, 2016). While most people experience feeling pain and causing pain to others as aversive, studies showed that individuals with high levels of psychopathic traits show atypical neural activity in response to imagining others' pain (Decety, Chen, Harenski, & Kiehl, 2013; Seara-Cardoso, Viding, Lickley, & Sebastian, 2015). Psychopathy is also associated with altered pain experience, which could (partly) explain their reduced empathic responses to others in pain. For instance, Marcoux and colleagues (2014) found a higher pain threshold in people with psychopathic tendencies. In addition, it was suggested that individuals with high levels of psychopathic traits attend to and experience pain differently (Van Heck et al. 2017; chapter 4). Furthermore, empathy processes are dependent on intact learning about stimulus-reinforcement and response-outcome associations, which also involve the amygdala. Disturbed learning of these associations results in inaccurate representations of the value of cues, actions or responses in the vmPFC. Thus, impairments in amygdala and the vmPFC functioning are proposed to underlie core impairments revelant to empathy-related processes in psychopathy.

Disturbances in associative learning and and outcome representation affect decision-making in psychopathy as well (Blair et al., 2004; Mitchell, Richell, Leonard, & Blair, 2006; Newman & Kosson, 1986; Patterson & Newman, 1993). Psychopathic individuals tend to make decisions that harm others (Mathieu & Babiak, 2016), and do not learn optimally about the negative consequences of these decisions for themselves and others. Particularly, striatal impairments reflect abnormal prediction error signaling, that is the signaling of differences between the amount of reward (or punishment) that is expected and the amount that is received (Blair, 2013). As a consequence, these individuals show a decreased ability to respond to changes in the environment, which is reflected in their impulsive and antisocial behavior. According to the IES model, impairments in the vmPFC and the striatum are suggested to underlie the core decision-making deficit in individuals with high levels of psychopathic traits (Blair, 2013).

1.7 Psychopathic traits and decision-making

The cognitive and affective deficits of individuals with psychopathic tendencies have a large impact on their choice behavior. Often these choices are made in a social context, thereby affecting the well-being of others in a negative way. An extensive amount of research has been devoted to understanding the processes underlying decision-making in individuals with psychopathic traits. However, the amount of proposed neurocognitive models is substantial and a single answer to the presumably simple question "Why do psychopathic individuals make poor choices?" remains elusive.

The IES model highlighted the core regions that seem implicated in psychopathy and provided an explanation for how these interacting neural systems give rise to the cognitive impairments that characterize psychopathy. Although this model, and other neurobiological and cognitive models, contributed significantly to our understanding of psychopathy, it also has its limitations. First, most of the individual core regions and impairments that were described by the IES are not unique to psychopathy. For instance, impairments in striatal functioning deficits have been associated with other populations that are characterized with impulsive and antisocial tendencies, such as ADHD (Jensen et al., 2001; Oldehinkel et al., 2016), externalizing (Byrne, Patrick, & Worthy, 2016; Gatzke-Kopp et al., 2009), and substance disorders (Yau et al., 2012; Tervo-Clemmens, Quach, Calabro, Foran, & Luna, 2020). Second, the direction of altered functioning in some of the regions and the associated impairments are not always found to be consistent. For instance, although most studies report reduced amygdala activation in relation to psychopathy (Birbaumer et al., 2005; Glenn, Raine, & Schug, 2009; Kiehl et al., 2001; Rilling et al., 2007), some findings suggest the opposite (Schultz & Balderton, 2016). Such inconsistencies could be (partly) explained by the fact that there are few studies that investigated the neurocognitive characterizations in relation to the subtypes that were mentioned by the personality-based conceptualizations of psychopathy. A more general criticism on the IES concerns its limited scope, due to the emphasis on the amygdala, vmPFC, and the striatum. A growing body of evidence indicates that the deficits linked to psychopathy are not confined to the amygdala and vmPFC (Anderson & Kiehl, 2012). Additional brain regions, prominently the anterior cingulate (Hornak et al., 2003), the insula (Weller, Levin, Shiv, & Bechara, 2009), and temporal pole (superiortemporal gyrus) (Ermer et al., 2012) were suggested to be implicated as well. Lastly, the overlap of diagnoses based on the core regions and impairments associated with psychopathy, as well as the inconsistencies in the literature, also demonstrate the lack of specificity of the neural regions and underlying mechanisms of psychopathy. The current neurocognitive models indicate the cognitive processes that are believed to be impaired and provide a broad interpretation of the mechanisms that give rise to these processes. However, cognitive processes have been suggested to consist of smaller interacting units that reflect aspects of cognition or brain function which are difficult to observe with our traditional experimental approaches (i.e. latent components). Neurocognitive models of psychopathy, including the IES, have been based on behavioral and neuroimaging findings, and therefore do not reflect the exact latent cognitive operations and interactive processes. Thus, they provide only a limited view on the computational processes that underlie behavior. The need for greater specification is also reflected by the current situation regarding treatment of antisocial offenders in general. Although a variety of treatment approaches have been described in literature, the responsivity to treatment in this group is relatively low. In addition, there is no compelling evidence for positive treatment outcomes for psychopathic offenders in particular (Harris & Rice, 2006). Deeper insight into the underlying impairments could be used to further stratify antisocial individuals and could facilitate the development of personalized treatment interventions that are tailored to fit the characteristics of these individuals.

Computational modelling is suggested as an alternative to study mechanistic properties on a deeper level (Kaplan & Craver, 2011; Mars, Shea, Kolling, & Rushworth, 2012). Such mathematical models provide a mechanistic account of what computations are performed and how they generate behavioral or neuro-physiological data. They allow us to zoom in on the exact sources underlying the cognitive impairments and investigate its function, hierarchical structure and interaction. To date, only a few studies have employed computational models that allow for a more detailed view into the latent cognitive operations underlying impairments in psychopathy (Blair, 2013; Brazil et al., 2013; Brazil, Mathys, Popma, Hoppenbrouwers, & Cohn, 2017; Oba, Katahira, & Ohira, 2019). These results provided important insights into reinforcement-learning and decision-making in individuals with psychopathic traits. Insight into the mechanistic account of psychopathy could open new avenues for developing treatment interventions targeting impairments in specific cognitive mechanisms.

1.8 Outline of the thesis

In this thesis, I aim to dive deeper into some of the mechanisms that are believed to play a role in the cognitive disruptions seen in individuals with high levels of psychopathic traits. More specifically, I combined behavioral, electrophysiological and computational approaches to study affective and social behavior in a series of experiments.

Chapter 2 describes an experiment in which distinct antisocial profiles in male offenders were investigated by performing latent profile analysis using the Self-Report Psychopathy Checklist Short Form (SRP-SF). This experiment yielded extensive and multifaceted characterizations of the different profiles that were previously suggested by studies using the PCL-R, thereby attributing to the validity of the SRP-SF. The latter was important for the other studies, as the SRP-SF was used to recruit participants based on their level of psychopathic traits.

The following chapters can be divided in two sections. The first part encompasses two chapters in which social-affective functioning is examined in relation to psychopathic traits. The IES model suggests that psychopathic behavior is driven by disturbances in specific associative learning processes of affective information in the amygdala. One particular aspect that seems to be impaired is the automatic processing of distress cues. Previous studies found support for this hypothesis in antisocial and psychopathic offender samples (Dawel, O'Kearney, McKone, & Palermo, 2012; Marsh & Blair, 2008). In **chapter 3,** automatic approach and avoidance responses to emotional facial expressions were assessed in a sample of healthy adults with a range of low to high levels of psychopathic traits. As a recent study suggested reduced threat avoidance in psychopathic offenders (von Borries et al., 2012), our main interest was the automatic response towards angry facial expressions. We were interested whether this effect was associated with the level of psychopathic traits in a community sample, and, more importantly, whether testosterone mediates this effect given its link with aggression and threat approach.

Another affective modality that has often been linked to psychopathy is the processing of pain. **Chapter 4** describes a study in which we examined pain sensitivity and empathy by measuring event-related potentials (ERPs) extracted from the ongoing EEG in an interactive setup. Each participant first fulfilled the role of "villain" (observing another person receiving electronic shocks) and later of "victim" (receiving the shocks while another person is watching). In addition, control over the painful stimulus was modulated, where "passive" refers to having no control over the shocks, while "active" refers to having control over delivering the shocks. This resulted in four different conditions; passive villain, active villain, active victim and passive victim. In addition, their level of psychopathic traits was measured. We were interested in sensitivity and processing of pain in self and others and how this is related to psychopathic traits.

Next to the affective impairments, the IES postulates that impaired decision-making is a core component of psychopathy. Nowadays, in our socially complex society, most of our decisions have an effect on others. High levels of psychopathic traits have been linked to poor social decisions and a disregard for the impact of own decisions on oneself and others. The second part of this thesis describes two studies that examined mechanisms of decision-making and the link to psychopathic traits. In **Chapter 5,** we investigated how individuals make use of social and non-social information in a reinforcement learning task in which the trade-off between the two types of information affects the task performance and the associated monetary reward for the participant. More specifically, we examined latent cognitive processes that are involved in associative learning of stable and volatile information and investigated the effect of psychopathic traits. In addition, we studied oscillatory theta activity given its potential involvement in adaptive control processes.

Chapter 6 describes an experiment in which moral strategies (i.e. decision styles) were studied by assessing reciprocity in an interactive socio-economic trust game. In such a context, choice behavior of the participant affects both the participant and the confederate. We applied a computational model to estimate the role of different moral strategies in this task and examined how this was related to psychopathic traits.

The final chapter, **Chapter 7**, I will provide a summary of the findings that were presented in each chapter, integrate the key findings, and discuss them in the light of the IES model. Furthermore, implications for future research will be discussed.

2

A comparison of latent profiles in antisocial male offenders

This chapter is based on:
Driessen, J. M. A., Fanti, K. A., Glennon, J. C., Neumann, C. S., Baskin-Sommers, A. S., & Brazil, I. A. (2018). A comparison of latent profiles in antisocial male offenders. *Journal of Criminal Justice, 57,* 47-55.

Abstract

The purpose of the present study is to identify distinct antisocial profiles in male offenders and investigate how these profiles differentiate based on psychopathic personality correlates and personality traits linked to criminogenic factors. Such profiles could provide a more complete view of the individual, and in the future, could aid diagnosis and foster the development of personalized treatment programs for individuals showing severe antisocial behaviour. First, we investigated the robustness and replicability of the profiles reported by previous profiling studies by performing latent profile analysis using the Self-Report Psychopathy Short-Form. Second, we studied how these profiles differentiate based on personality correlates believed to be relevant to psychopathy. Third, we investigated how each profile relates to externalizing behavior. Four antisocial profiles were identified: *generic offenders*, *impulsive-antisocial traits offenders*, *non-antisocial psychopathic traits offenders*, and *psychopathic traits offenders*. The validity of these profiles was supported by their links with external variables concerning psychopathic personality correlates and externalizing behaviors. Consistent with previous research using the PCL-R, the present study provides support for the presence of four distinct antisocial profiles based on self-report psychopathy scores in a male offender sample. Furthermore, findings provide relatively extensive and multifaceted characterizations of each profile.

2.1 Introduction

Antisocial behaviour is a heterogeneous construct that covers a wide range of behaviours that cause harm to others. Importantly, it is now recognized that these behaviours represent the outcome of different etiological pathways (Baskin-Sommers et al., 2015), and there is evidence supporting the existence of different subtypes of antisocial individuals (for an overview see Brazil, van Dongen, Maes, Mars, & Baskin-Sommers, 2016), which seem to differ in their externalizing tendencies (Patrick et al., 2005) and the level of violence involved in their crimes (DeLisi et al., 2011; Odgers et al., 2007; Vincent, Vitacco, Grisso, & Corrado, 2003). Several taxonomies have been proposed that differ in how the subtypes are characterized, but most of these views share the notion that there is a distinction between antisocial individuals with and without psychopathy (Brazil et al., 2016; Kiehl & Hoffman, 2011). Traditionally, psychopathy has been seen as a severe disorder typified by interpersonal-affective dysfunctions (e.g., lack of empathy, manipulativeness) combined with severe antisocial behaviour and an erratic lifestyle. Psychopathy has been linked to increased chance of recidivism (McCuish, Corrado, Hart, & DeLisi, 2015), excessive use of aggression, and large financial costs to society (Kiehl & Hoffman, 2011).

The currently dominant approach to differentiate between psychopathic- and non-psychopathic antisocial individuals is based on the framework developed by Hare and colleagues (1980). Driven by the idea that there was no appropriate measure to diagnose antisocial individuals at the time, Hare developed the Psychopathy Checklist (Hare, 1980), and later on the Psychopathy Checklist-Revised (PCL-R; Hare, 2003; Hare et al., 1990). The PCL-R is a semi-structured interview that can be combined with criminal records to derive a score that indicates the extent to which psychopathic characteristics are present in an individual. An individual is diagnosed with psychopathy if the total score of the PCL-R>30 in the U.S., or >26 in Europe (Cooke & Michie, 1999). However, the PCL-R score represents the combination of four dimensions or facets believed to constitute psychopathy. The 'interpersonal' facet concerns arrogant and deceitful interpersonal style, which is characterized by superficial charm, grandiosity, manipulative behaviour and deceitfulness. The 'affective' facet captures the degree of disturbed affective experience, which encompasses callousness, lack of empathy, failure to accept responsibility and lack of remorse or guilt. The 'lifestyle' facet describes an impulsive-irresponsible behavioural style, which is typified by impulsivity, boredom, sensation seeking, a parasitic lifestyle, irresponsibility, and lack of goals. Finally, the 'antisocial' facet encompasses aggressiveness, early behaviour problems, juvenile delinquency and criminal versatility (Hare & Neumann, 2005). These facets are inter-related and load on a set of second-order factors, forming an Interpersonal-Affective Factor (Factor 1; F1) and a Lifestyle-Antisocial Factor (Factor 2; F2). Whereas the Interpersonal-Affective factor captures the core features that are unique to psychopathy, the Lifestyle-Antisocial factor represents a more general set of antisocial tendencies that can be found across several subtypes of antisocial individuals (Hansen et al., 2007; Hare, 2003). The framework offered by the PCL-R has received extensive empirical support in a wide variety of samples and is now regarded as the most reliable method to measure psychopathic traits (Hare, Clark, Grann, & Thornton, 2000; Hare & Neumann, 2006; Neumann, Schmitt, Carter, Embley, & Hare, 2012).

Historically, there have been many proposals on how to subtype individuals based on psychopathic features (see Brazil et al., 2016). One of the most prominent distinctions has been that between primary and secondary psychopathy, which has been defined in various ways. For instance, primary psychopathy has been described as antisocial individuals that score relatively high on F1 traits compared to F2 traits, whereas secondary psychopathy has been characterized by relatively high F2 traits relative to F1 traits (Skeem et al., 2003; Wong & Hare, 2006). Others have suggested subtypes based on levels of anxiety. Primary psychopathy has been defined as a high PCL(-R) total score and a low level of anxiety, whereas secondary psychopathy has been characterized as a high PCL(-R) total score with a high level of anxiety (Lykken, 1995; Skeem, Johansson, Andershed, Kerr, & Louden, 2007). Another proposal is based on differences in behavioural motivation, and suggest that primary psychopathy is typified by an underactive behavioural inhibition system (BIS) in those scoring above the PCL-R cutoff score, while secondary psychopathy concerns an overactive behavioural activation system in these individuals (BAS; Newman et al., 2005; Ross et al., 2007). Furthermore, studies that aimed to subtype psychopathy suggested a role for emotional aspects (Hicks, Markon, Patrick, Krueger, & Newman, 2004; Hicks & Patrick, 2006) and externalizing behaviour (Patrick et al., 2005), where F1 traits were negatively associated with low negative emotionality and low externalizing behaviours, and F2 traits were positively associated with high negative emotionality and high externalizing behaviours. Importantly, however, many proposals on how to best subtype individuals with psychopathy were based on theoretical assumptions (e.g., Murphy & Vess, 2003; Skeem et al., 2003), and often use diverse methodologies, as well as different sample selection procedures (Neumann, Vitacco, & Mokros, 2016).

Recent studies have employed structural equation modeling as a quantitative approach to subtyping of antisocial individuals. For example, Skeem et al. (2007) performed a model-based cluster analysis on a sample of Swedish male offenders with a PCL-R score >28. The clustering was based on the four PCL-R facet scores and a self-report measure of trait anxiety. The analysis resulted in two clusters

with one type (60% of the sample) scoring high on PCL facets 1-3 (interpersonal, affective, lifestyle), but low on anxiety and the other type (40% of the sample) showing a moderate score on PCL facets 1-3 and high on anxiety. Notably, the antisocial facet did not differ between the two clusters. A more recent study by Mokros and colleagues (2015) used Latent Class Analysis (LCA) on PCL-R data from male offenders with a high PCL-R score (>27). Three subtypes were obtained: manipulative (Latent Class 1), aggressive (Latent Class 2) and sociopath (Latent Class 3). The manipulative and aggressive classes reflected early clinical conceptualizations of psychopathy and were proposed to represent empirically derived variants of primary psychopathy that differ in the manifestation of F1 and F2 traits. Moreover, the sociopath class was believed to reflect secondary psychopathy as this latent class was characterized by social deviance, and low expression of the affective features of psychopathy. Whereas previous studies were mostly conducted using offender samples with high PCL-R scores, some recent studies have examined the full range of PCL-R scores in mixed offender (Hare, 2016) and sex offender (Krstic et al., 2017) samples. These studies provided evidence for the existence of four latent classes: psychopaths, callous-conning offenders, sociopaths and general offenders. The general offenders were at the low end of the psychopathic spectrum, and the psychopaths were at the high end of the spectrum. The sociopaths showed mainly elevated F2 traits, while elevated F1 traits were the most prominent features of the callous-conning offenders. Taken together, findings from these studies suggest that antisocial behaviour can be subtyped by using psychopathy measures, and these subtypes represent different profiles with regard to psychopathic traits.

Notably, the vast majority of the current empirical research on subtyping of psychopathy has predominantly been based on the PCL(-R). However, administering and scoring the PCL-R requires a relatively large time investment. Therefore, self-report measures of psychopathy are gaining popularity in forensic research, especially in studies that are interested in subtyping psychopathic traits in the general population (e.g., Colins, Fanti, Salekin, & Andershed, 2017). The Self-Report Psychopathy checklist (SRP; Hare, 1985) is a well-known self-report questionnaire for psychopathic traits which uses a similar four-dimensional structure to the PCL-R. The SRP is significantly associated with the PCL-R (latent r=0.68) and has been proven to be valid across genders (Neumann & Hare, 2008; Neumann & Pardini, 2014). However, to date, there are no studies addressing the suitability of self-report measures for subtyping of adult offenders based on psychopathic features.

The main purpose of the present study was to identify different antisocial profiles in a sample of male offenders and investigate how these profiles differ based on general personality factors and other traits linked to criminogenic factors. To achieve this, we (1) performed latent profile analysis (LPA) on the SRP-Short Form and compared our results with the only three previous studies that employed LPA in adult offenders, (2) studied how the profiles differed on descriptive and personality factors traditionally believed to be relevant for distinguishing among subtypes based on levels of psychopathy (e.g., anxiety, valence of affect, motivational tendencies), and (3) sought to further extend previous studies on subgrouping in adult offenders by also obtaining a more detailed view of how the profiles differed on externalizing behaviors commonly seen in antisocial offender populations (i.e., aggression, disinhibition, substance abuse). LPA is a data-driven approach that classifies individuals or cases into homogenous groups (i.e., latent profiles) based on conditional probabilities. This is in contrast with the majority of the subtyping studies that have used hypothesis-driven analyses and are dependent on strong a priori assumptions. Given that there have only been three studies examining PCL-based subtypes of offenders using LCA in incarcerated adult offenders and that these studies differed in the number of latent classes identified, it is important to test whether these results are stable and replicate when using alternative measures of psychopathy derived from the PCL-R. Based on previous literature, we expect to identify four different profiles that are similar to previously identified subtypes. Regarding the external variables, we expect that the profile with the lowest SRP scores is characterized by high anxiety score. Furthermore, we expect that the profile with high scores on antisocial behavior is associated with high rates of violent crimes. Also, we expect that the profile with the highest SRP scores is linked to a high amount of crimes, low anxiety scores and high negative affectivity and reward sensitivity.

2.2 Methods

2.2.1 Participants

The current study investigated behavioural data from 576 male offenders that were institutionalized in maximum security prisons throughout Wisconsin. Participants ranged in age from 18 to 45 years (*M*=31.31, *SD*=7.13). Race among the sample was heterogeneous, with 389 participants (67.5%) self-identifying as White, 142 participants (24.7%) as Black, 17 (3.0%) as Hispanic, 3 (0.5%) as Native American, one (0.2%) as Asian and two (0.3%) as a mix of two or more races. Procedures were approved by the Ethics Committee of the University of Wisconsin (IRB SE-2011-0358).

2.2.2 Questionnaires

Psychopathy. Psychopathic traits were assessed with the short form of the Self-Report Psychopathy scale (SRP-SF; Paulhus, Neumann & Hare, 2015). This measure was designed to assess the four facets of psychopathy as described by Hare's PCL-R framework; interpersonal manipulation (e.g., "Sometimes you need to pretend that you like someone to get what you want"), affective callousness (e.g., "I never feel guilty over hurting others"), erratic lifestyle (e.g., "I've often done dangerous things just for the thrill") and overt antisociality (e.g., "Sometimes I carry a weapon (knife or gun) to protect myself"). A total of 29 questions were rated on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). The full version SRP is strongly correlated (r=0.92) with the SRP-SF (Paulhus et al., 2015) and the PCL-R (Declercq, Carter, & Neumann, 2015). The SRP and SRP-SF both have good basic psychometric properties (Neumann et al., 2012), are theoretically sound (Carré, Fisher, Manuck, & Hariri, 2012), and have robust latent structures (Neumann & Pardini, 2014; Williams, Paulhus, & Hare, 2007). Internal consistency in our sample was high for the total score (Cronbach's α=.88), and acceptable for the factor scores (interpersonal α=.80, affective α=.67, lifestyle α=.70, antisocial α=.63).

External correlates. A set of measures was selected as external correlates due to their relevance for subtyping of psychopathy. **Anxiety** was measured with the Welsh Anxiety Scale (WAS; Welsh, 1952), which consists of 39 items that are rated on a true/false scale. **Motivational tendencies** were measured with the Behavioural Inhibition/ Behavioural Activation scales (BIS/BAS; Carver & White, 1994). This measure consists of 20 items measuring four scales: BIS, BAS – Reward Responsiveness, BAS – Drive, and BAS – Fun Seeking. The items were rated on a 4-point Likert scale (1=strongly agree, 4=strongly disagree) (Jorm et al., 1998). Where the BIS scale is related to neuroticism and negative affect, BAS is more related to extraversion and positive affect (Carver & White, 1994). The three BAS subscales load strongly on the second-order BAS scale. Therefore, we only included the higher-order BAS scale in our analysis. **General emotionality** was assessed with two subscales of the Multidimensional Personality Questionnaire (MPQ; Tellegen & Waller, 2008). The brief version of this questionnaire consists of 155 items that measure three subscales: Positive Emotionality; Negative Emotionality and Constraint. The Positive Emotionality subscale consists of four lower-order scales (Well-Being, Social Potency, Achievement, and Social Closeness). Negative Emotionality has three lower-order factors (Stress Reaction, Alienation, and Aggression), as does Constraint (Control, Harm Avoidance, and Traditionalism). The Constraint subscale was not taken into account in the current analysis. **Externalizing behaviour** was measured with a brief version of the Externalizing Spectrum Inventory (ESI; Krueger, Markon, Patrick, Benning, & Kramer, 2007) in a subgroup (N=355) of the total sample. The ESI is a well validated self-report questionnaire that is used to measure disinhibitory behaviours and traits in both clinical and research settings (Venables & Patrick, 2012; Widiger & Sankis, 2000). The brief version used in the current study encompasses 100 items that form 18 subfactors, which ultimately measures three superordinate factors: Inhibition, Aggression and Substance abuse.

Descriptive variables. Intelligence and number of offenses were also analyzed given prior indications that antisocial subgroups could differ on these variables (Kandel et al., 1988; Laurell, Belfrage, & Hellström, 2010). Intelligence was measured with the Wechsler Adult Intelligence Scale IV (WAIS-IV; Wechsler, 2008), which is considered to be one of the best measures of general intellectual functioning. Offending was assessed using the number of violent and non-violent crimes convictions.

2.2.3 Statistical analyses

Latent Profile Analysis (LPA). In the present study, we conducted LPA, in which the variables used to generate classes (or profiles) are continuous instead of categorical. LPA is an extension of Latent Class Analysis, and estimates the probability of an individual or case to belong in one class versus another class based on a set of observable characteristics. One of the assumptions of LPA is that the variables are independent within each latent class, as correlations between variables are explained through the structure of the classes. Statistical criteria are used to determine the number of classes that best describe the data (Bauer & Curran, 2004).

Mplus 7 (Muthén & Muthén, 1998-2015) was used to run the LPA and identify subtypes in our sample (*N*=567) based on SRP-SF scores. The four SRP facets interpersonal (Int), affective (Aff), lifestyle (Lif), and antisocial (Ant), were used as observed variables for the LPA model. The optimal number of classes was defined based on a set of statistical criteria: the Bayesian information criterion (BIC), the Lo-Medel-Rubin (LMR) statistic, the posterior probabilities, and the entropy value (Nylund, Asparouhov, & Muthén, 2007). The model with the lowest BIC value was considered the best model. The LMR is considered as a likelihood ratio test between models with a different number of latent classes. It tests k-1 classes compared to k classes and results in a chi-square value that indicates whether the k-1 class model should be rejected in favor of the k class model. The posterior probabilities are considered to determine the accuracy of the classification and the entropy value (ranges from zero to one) gives an indication of the amount of diversity of the latent classes. Both posterior probability and entropy imply satisfactory fit when values exceed .70 (Muthén, 2000; Nagin, 2005). Because LPA might result in some individuals being misclassified (Bakk, Tekle, & Vermunt, 2013), scores of all participants were inspected and classification errors where

corrected when necessary. After correction, the model was retested and the fit measures based on the corrected model are reported.

Pairwise comparisons. Following the LPA, questionnaires were imported to the statistical analysis program JASP (JASP Team, 2016, Version 0.8.0.0) and the profiles were compared on several personality and behavioural measures. A one-way Analysis of Variance (ANOVA) was conducted to test for significant differences (α=0.05) between item averages of the latent classes. Post-hoc comparisons were further conducted to obtain all possible pairwise comparisons between the profiles. Standard Bonferroni correction was used to account for multiple comparisons.

In addition, we repeated the analyses using Bayesian independent t-tests and calculated Bayes Factors (BF) to determine how likely to be true the results obtained with each group comparison were, given the data. Some advantages of using Bayesian statistics are that it can provide a quantification of the evidence *supporting* of the null-hypothesis, rather than only against it (Wetzels, Odekerken-Schröder, & Van Oppen, 2009), and this statistical approach does not suffer from the drawbacks of classical testing, such as the need to correct for multiple comparisons and reliance on various assumptions (Morey, Rouder, Verhagen, & Wagenmakers, 2014). We considered an effect to be strongly supported only when both analytical approaches yielded a similar result.

2.3 Results

2.3.1 Latent Profile Analysis

To identify the optimal number of groups to retain, models with one to five classes were estimated using LPA. To obtain the best possible solution, we repeated this procedure by using automatic, random, and user-specified starting values, and by relaxing the default equality constraints used in Mplus (i.e., means and variances of the latent class indicators). The BIC statistic increased from Class 4 (BIC=13235.01) to Class 5 (BIC=13247.89) and decreased from Class 3 (BIC=13279.73) to Class 4. In addition, the LMR statistic fell out of significance for the five-class model (*p*=.26). Thus, the four-class model better represented the data based on the BIC and LMR statistics. The mean posterior probability scores ranged from .85 to .90 and the entropy value was .76, suggesting that the identified classes were well separated. The four-class model included a group referred to as the *generic offenders* of 256 male offenders whom scored low on all facets of the SRP (Int: *M(SD)*=13.15(3.15), Aff: *M(SD)*=14.53(2.76), Lif: *M(SD)*=17.28(4.30), Ant: *M(SD)*=16.33(3.28)). The second group of 89 male offenders, labeled as *impulsive-antisocial traits offenders*, scored high on the lifestyle and antisocial facet of the SRP and relatively low on two other facets of the SRP (Int: *M(SD)*=16.77(3.26), Aff: *M(SD)*=17.17(2.69), Lif:

M(SD)=21.61(4.36), Ant: *M(SD)*=25.99(2.64)). The third group, referred to as the *non-antisocial psychopathic traits offenders*, included 121 male offenders whom scored low on the antisocial facet of the SRP and relatively high on the other three facets of the SRP (Int: *M(SD)*=19.80(3.87), Aff: *M(SD)*=20.15(3.03), Lif: *M(SD)*= 22.62(3.95), Ant: *M(SD)*=19.44(3.07)). The last group was labeled as *psychopathic traits offenders* and consisted of 101 male offenders whom scored high on all facets of the SRP (Int: *M(SD)*=23.01(4.20), Aff: *M(SD)*=23.00(3.06), Lif: *M(SD)*= 25.31(3.93), Ant: *M(SD)*= 27.75(3.53)). Additional analyses indicated that there were no differences across groups based on age ($\chi^2(81)=99.67$, $p=.08$) and race (white vs. others: $\chi^2(15)=14.66$, $p=.48$). An overview of the latent classes is shown in Figure 2.1. Means and standard deviations of the psychopathic subscales for each profile are represented in Table 2.1 (for BFs see Table S2.1 in the Supplementary materials).

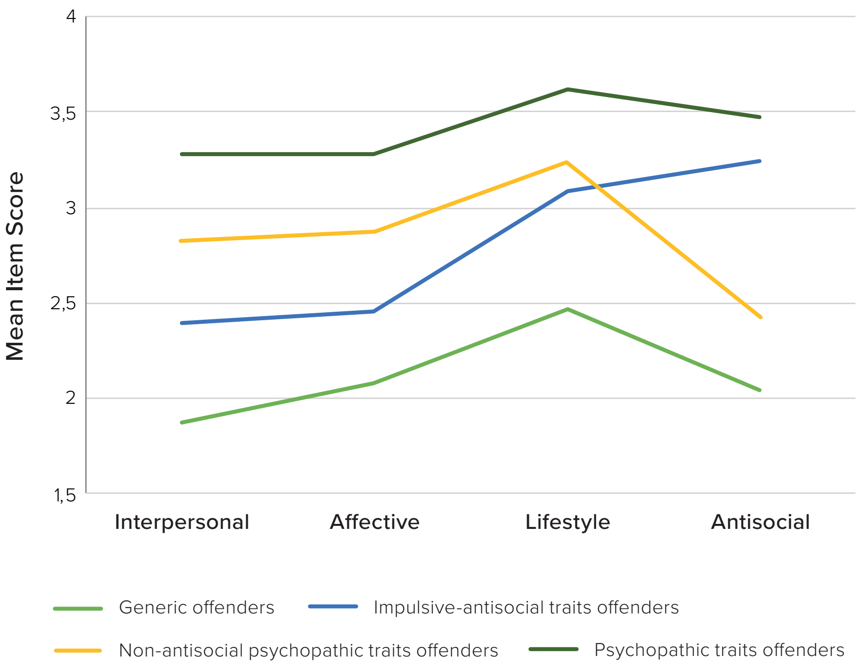


Figure 2.1 Latent profiles of antisocial behaviour in male offenders. Mean item score of each latent profile on the SRP factors.

Table 2.1 Means and standard deviations of the SRP facet scores and external variables for each profile

	GE	IA	NA	PS		
	M(SD)	M(SD)	M(SD)	M(SD)	F	partial η^2
SRP Interpersonal	13.15^{a} (3.15)	16.77^{b} (3.26)	19.80^{c} (3.87)	23.01^{d} (4.20)	228.09	0.54
SRP Affective	14.53^{a} (2.76)	17.17^{b} (2.69)	20.15^{c} (3.03)	23.00^{d} (3.06)	253.53	0.57
SRP Lifestyle	17.28^{a} (4.30)	21.61^{b} (4.36)	22.62^{b} (3.95)	25.31^{c} (3.93)	109.78	0.37
SRP Antisocial	16.33^{a} (3.28)	25.99^{c} (2.64)	19.44^{b} (3.07)	27.75^{d} (3.53)	416.72	0.69
WAIS IQ	98.40 (12.13)	97.54 (11.59)	98.94 (11.58)	98.85 (12.37)	0.25	0.00
APD	5.88^{a} (3.04)	10.22^{c} (3.46)	8.25^{b} (3.67)	11.37^{c} (3.63)	79.75	0.30
Violent crimes	3.35^{a} (3.43)	7.04^{c} (8.80)	3.37^{ab} (3.36)	4.87^{b} (3.92)	14.72	0.09
Non-violent crimes	9.61^{a} (10.67)	15.01^{b} (15.70)	14.03^{b} (10.64)	15.49^{b} (15.13)	8.28	0.05
Welsh-anxiety	9.84^{a} (7.85)	14.59^{b} (8.95)	14.18^{b} (8.17)	16.80^{b} (8.99)	20.79	0.10
MPQ-positive	65.24 (14.62)	66.22 (13.40)	66.27 (15.99)	67.59 (15.73)	0.60	0.00
MPQ-negative	39.11^{a} (15.88)	51.92^{b} (15.95)	57.53^{c} (16.59)	67.40^{d} (15.29)	87.20	0.34
BIS	18.46^{b} (3.29)	18.39^{ab} (3.57)	17.54^{ab} (3.83)	17.25^{a} (3.74)	3.81	0.02
BAS	36.86^{a} (4.71)	38.10^{a} (5.05)	39.89^{b} (5.19)	41.68^{c} (5.50)	24.78	0.13

Notes. Means with different superscripts differ significantly from each other ($p<0.05$). GE = generic offenders, IA = impulsive-antisocial traits offenders, NA=non-antisocial psychopathic traits offenders, PS = psychopathic traits offenders.

2

2.3.2 Descriptives and personality correlates of the latent profiles

Means were compared across the latent profiles (see Table 2.1, for BFs Table S2.1). **Anxiety** measures showed that the *generic offenders* scored significantly lower on anxiety than the other profiles. **Negative emotionality** differed significantly across all profiles. The *generic offenders* reported lowest negative emotionality, followed by the *impulsive-antisocial offenders,* the *non-antisocial psychopathic traits offenders* and the *psychopathic traits offenders.* **Motivational tendencies** were measured with BIS and BAS. The *generic offenders* scored significantly higher on BIS than the *psychopathic traits offenders.* The *impulsive-antisocial traits offenders* and *non-antisocial psychopathic traits offenders* did not show any significant differences in motivational tendencies compared to the other profiles. The BAS score was significantly higher in the *psychopathic traits offenders* compared to the other profiles. The *non-antisocial psychopathic traits offenders* reported a significantly higher BAS score than the *generic offenders* and the *impulsive-antisocial traits offenders.* **Offending** was measured as violent- and non-violent crime rates. The *non-antisocial traits offenders* and the *psychopathic traits offenders* reported a significantly higher number of violent crimes compared to the other profiles. The *non-antisocial psychopathic traits offenders* reported a significantly higher number of violent crimes compared to the *generic offenders.* The latter profile also reported a lower number of non-violent crimes compared to the other profiles. **Intelligence** and **positive emotionality** scores did not differ among profiles. The scores on the external variables for each of the latent classes are visualized in Figure 2.2.

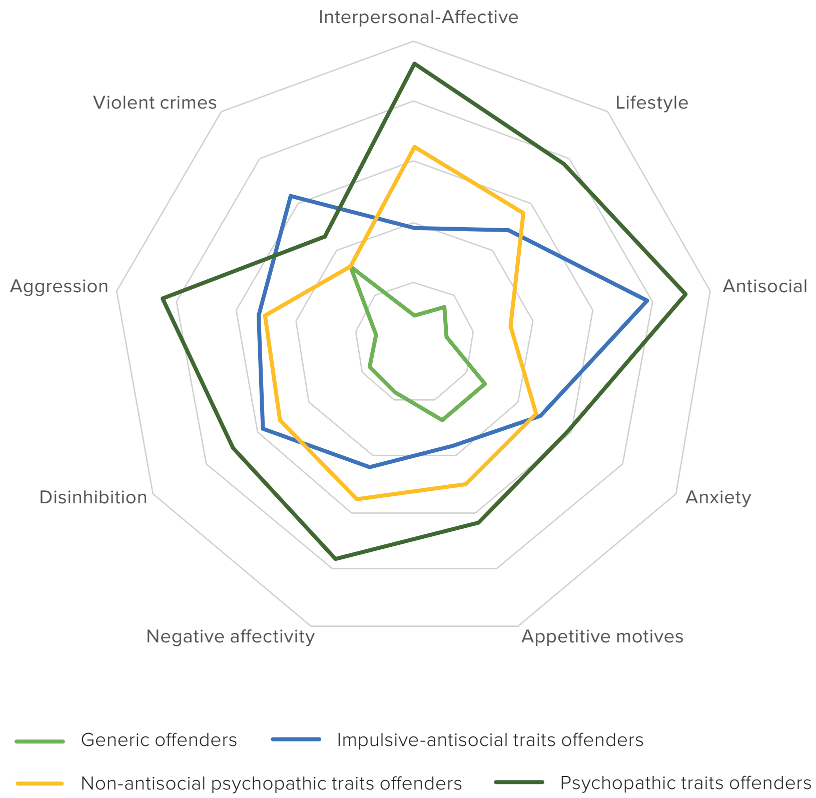


Figure 2.2 Psychopathic and external correlates of the latent profiles. Mean z-scores of each latent profile on the SRP factors and a selection of the external correlates. Interpersonal-affective, lifestyle, antisocial (SRP), anxiety (Welsh anxiety), appetitive motives (BISBAS), negative affectivity (MPQ), disinhibition, aggression (ESI), and violent crimes.

A representative subset of the total sample (*N*=355; *generic offenders*=161, *impulsive-antisocial psychopathic traits offenders*=57, *non-antisocial psychopathic traits offenders*=73, *psychopathic traits offenders*=64) filled in the ESI to measure **externalizing behaviours (**see Table 2.2, for BFs see Table S2.2). Eighteen subscales were represented by three higher-order factors: Disinhibition, Callous-Aggression, and Substance Abuse. The *psychopathic traits offenders* scored significantly higher on disinhibition compared to the *non-antisocial psychopathic traits offenders* and the *generic offenders.* The *impulsive-antisocial traits offenders* scored significantly higher than the *generic offenders.* The latter profile scored significantly lower on Disinhibition compared to all other profiles. The level of callous aggression was the highest in the *psychopathic traits offenders,* while the *generic offenders* scored lower compared to the other profiles. The *generic offenders* scored significantly lower on substance abuse compared to the *impulsive-antisocial traits offenders* and the *psychopathic traits offenders.*

2.4 Discussion

The main purpose of the present study was to identify and characterize different antisocial subtypes in a general male offender sample based on a self-report measure of psychopathic traits by using latent profile analysis. Four antisocial profiles were identified: *generic offenders*, *impulsive-antisocial traits offenders*, *non-antisocial psychopathic traits offenders*, and the *psychopathic traits offenders*. The *generic offenders* scored relatively low on all dimensions of the SRP. The *impulsive-antisocial traits offenders* scored high on the antisocial and lifestyle facets, whereas the other dimensions were relatively low. The *non-antisocial psychopathic traits offenders* showed relatively high scores on the interpersonal, affective and

Table 2.2 Means and standard deviations of the externalizing measures for each profile.

	GE	IA	NA	PS		
	M(SD)	M(SD)	M(SD)	M(SD)	F	partial η^2
ESI-Disinhibition	2.23^{a} (0.57)	2.99^{bc} (0.45)	2.84^{b} (0.50)	3.14^{c} (0.48)	49.52	0.3
Irresponsibility	2.17^{a} (0.67)	2.72^{b} (0.62)	2.57^{b} (0.67)	2.71^{b} (0.72)	16.01	0.12
Impatient urgency	2.40^{a} (1.10)	3.19^{b} (0.99)	3.24^{b} (0.83)	3.48^{b} (0.69)	25.72	0.18
Problematic impulsivity	2.40^{a} (0.68)	3.14^{bc} (0.55)	2.89^{b} (0.60)	3.25^{c} (0.63)	35.97	0.24
Fraud	1.61^{a} (0.61)	2.44^{bc} (0.86)	2.22^{b} (0.65)	2.58^{c} (0.73)	42.23	0.27
Boredom proneness	2.76^{a} (1.05)	3.14^{ab} (1.03)	3.08^{ab} (0.97)	3.45^{b} (0.77)	8.34	0.07
ESI-Callous Aggression	1.71^{a} (0.46)	2.35^{b} (0.53)	2.32^{b} (0.47)	2.87^{c} (0.50)	95.97	0.45
Low empathy	1.24^{a} (0.50)	1.44^{a} (0.69)	1.79^{b} (0.77)	2.40^{c} (0.93)	46.55	0.29
Relational aggression	1.64^{a} (0.55)	2.12^{b} (0.70)	2.20^{b} (0.58)	2.64^{c} (0.74)	42.2	0.27
Excitement seeking	1.80^{a} (0.92)	2.18^{ab} (1.11)	2.37^{b} (0.99)	2.83^{c} (0.87)	19.59	0.15
Destructive aggression	1.63^{a} (0.65)	2.30^{b} (0.77)	2.05^{b} (0.70)	2.64^{c} (0.82)	33.04	0.22
Physical aggression	1.81^{a} (0.71)	2.87^{c} (0.65)	2.48^{b} (0.72)	3.18^{d} (0.62)	73.79	0.39
Honesty	1.86^{a} (1.02)	2.46^{b} (1.21)	2.43^{b} (1.17)	2.83^{b} (1.15)	13.41	0.1
Rebelliousness	2.01^{a} (0.93)	3.11^{b} (0.99)	2.93^{b} (0.92)	3.61^{c} (0.58)	58.49	0.34
ESI-Substance abuse	6.11^{a} (2.30)	7.20^{b} (2.34)	6.58^{ab} (2.25)	7.17^{b} (2.36)	4.39	0.04
Alcohol problems	16.50 (7.95)	18.98 (8.14)	17.60 (7.80)	19.20 (8.14)	1.99	0.02
Drug use	2.46^{a} (0.97)	3.17^{b} (0.85)	2.84^{b} (0.91)	3.14^{b} (0.97)	11.91	0.09
Marijuana problems	1.91^{a} (1.24)	2.70^{b} (1.75)	2.08^{ab} (1.22)	2.48^{b} (1.39)	6.64	0.05
Marijuana use	3.56^{a} (0.97)	3.94^{b} (0.29)	3.77^{ab} (0.74)	3.84^{ab} (0.54)	3.89	0.03

Notes. Means with different superscripts differ significantly from each other ($p<0.05$).

lifestyle dimensions, whereas the score on the antisocial facet was low. In contrast to the *generic offenders*, the *psychopathic traits offenders* showed high scores on all dimensions of the SRP.

2.4.1 LPA results based on the SRP-SF

Our first aim was to investigate the robustness of the previously reported findings on PCL-based profiles obtained using LCA by replication of this method while using the SRP-SF. Overall, our LPA outcomes resemble prior findings obtained in a large North American male offender sample (N=4865) using the PCL-R (Hare, 2016). The *impulsive-antisocial traits offenders* and the *non-antisocial psychopathic traits offenders* in our sample seem to parallel the primary (high F1) and secondary (high F2) psychopaths in Hare's sample. Despite the similarities, there were some differences with respect to the lifestyle facet. Hare (2016) and Krstic et al. (2017) found a significant difference on the lifestyle dimension between the primary and the secondary group, while our results showed similar scores on the lifestyle facet for the *impulsive-antisocial traits offenders* and the *non-antisocial psychopathic traits offenders*. One possible explanation could be that offenders under-report tendencies captured by the lifestyle facet, which is one of the potential risks of using self-report in general (Hindelang, Hirschi, & Weis, 1981; Van de Mortel, 2008). Still, and more importantly, the profiles identified in our study were differentially related to external variables proposed to be important for subtyping based on psychopathic features.

2.4.2 Psychopathic personality correlates

The second aim was to investigate possible group differences in personality traits proposed to be relevant to psychopathy between the profiles. Several studies have assigned a key role to anxiety in discriminating among different types of individuals with psychopathic traits (Burns, Roberts, Egan, & Kane, 2015; Lykken, 1995; Salihovic, Kerr, & Stattin, 2014; Skeem et al., 2007). The results showed that the *generic offenders* had significantly lower levels of anxiety compared to the other groups, but that the other groups did not differ from each other. That is, anxiety levels do not seem to differentiate well among subtypes of antisocial offenders in our sample, and our sample of *psychopathic traits offenders* included high-anxious individuals with psychopathic traits. At first this may seem to counter the general belief that psychopathy should be related to reduced trait anxiety. However, the possibility that anxiety plays a limited role in defining psychopathy has been highlighted before (Neumann & Hare, 2008; Neumann, Johansson, & Hare, 2013), and the results of a recent meta-analysis have also pointed out that anxiety is weakly related to psychopathy (Decuyper, De Pauw, De Fruyt, De Bolle, & De Clercq, 2009). Note, however, that we did not analyze the role of anxiety within the group of *psychopathic traits offenders*, so there is still a possibility that anxiety scores differentiate among subtypes *within* the psychopathic traits group (Lykken, 1995; Skeem et al., 2007).

Another factor proposed to have discriminative power in populations with psychopathic tendencies is behavioural motivation (Book & Quinsey, 2004; Newman et al., 2005). An underactive BIS has been suggested to explain a lack of fear for punishment in psychopathy (Gray, 1970; Patrick, 1994), while an overactive BAS has been linked to the hypersensitivity to reward often seen in psychopathy (Arnett, Smith, & Newman, 1997; Gorenstein & Newman, 1980). Consistent with this previous research, our findings point out that *psychopathic traits offenders* have an underactive BIS and overactive BAS compared to *generic offenders* (Hoppenbrouwers, Neumann, Lewis, & Johansson, 2015). Furthermore, *non-antisocial psychopathic traits offenders* showed an overactive BAS, while scores on the BIS were not different compared to other groups. The lifestyle dimension, which covers impulsive behaviour, was high in the *non-antisocial psychopathic traits offenders*. As high impulsivity rates are associated with a strong BAS (Franken, Muris, & Rassin, 2005), this finding was in the expected direction. The results for the *impulsive-antisocial traits offenders* indicate that the behavioural motivational system seems to play a less prominent role in differentiating these individuals from the other groups.

Negative emotional tendencies form a third factor that has been proposed to discriminate among subtypes based on psychopathy scores (Hicks et al., 2004; Hicks & Patrick, 2006). Our results pointed out that negative emotionality was low in the *generic offenders* and high in the *psychopathic traits offenders*. On one hand, this is in agreement with the suggestion that higher F2 traits are linked to increased negative emotionality in *psychopathic traits offenders* (Hicks et al., 2004; Hicks & Patrick, 2006), on the other hand this group also showed high F1 traits. However, as mentioned earlier, the *psychopathic traits offenders* seem to mainly include high-anxious psychopathic individuals as their anxiety score is higher compared to the *generic offenders*. This could be an explanation for the higher negative emotionality scores in the *psychopathic traits offenders*. A small effect was found when comparing the *impulsive-antisocial traits offenders* with the *non-antisocial psychopathic traits offenders*, in which the latter group showed higher scores on negative emotionality. This could seem counterintuitive, because the *non-antisocial psychopathic traits offenders* show more F1 traits compared to the *impulsive-antisocial traits offenders,* and increased F1 traits have been linked to decreased negative emotionality (Hicks & Patrick, 2006). However, this finding converges with the suggestions that non-psychopathic antisociality is linked to increased prevalence of conditions characterized by elevated negative affect, such as anxiety and mood disorders (Blair, 2012; Verona, Patrick, & Joiner, 2001).

2.4.3 Externalizing behavior and criminogenic factors

Externalizing was also proposed to be an important factor that is supposed to differ among subgroups of antisocial individuals (Patrick et al., 2005). Therefore, our third aim was to obtain a more detailed perspective on how each profile relates to externalizing behaviours commonly seen in antisocial offender populations. The *impulsive-antisocial traits offenders* and the *non-antisocial psychopathic traits offenders* showed a similar profile regarding externalizing, although the *impulsive-antisocial traits offenders* showed more physical aggression. In general, the *generic offenders* showed less externalizing compared to the other groups, especially relative to the *psychopathic traits offenders.* Similar to negative emotionality, high externalizing in the *psychopathic traits offenders* is suggested to be related to the elevated levels of F2 traits found in this latent profile (Patrick et al., 2005). Compared to all other groups, the *psychopathic traits offenders* showed significantly higher scores on the disinhibition and callous-aggression subscales of the ESI. While it is unsurprising that the *psychopathic traits offenders* generally show more externalizing behaviour than the other groups, we had not expected aggression-related tendencies to be lower in the *impulsive-antisocial traits offenders* compared to the *psychopathic traits offenders*. However, the ESI measures callous aggression, which is suggested to predict elevated levels of psychopathy and non-psychopathic antisociality, but in a lesser extent (Venables & Patrick, 2012). With regard to the subfactor substance use, there was no clear pattern of differences among the groups, which could be due to the fact that substance abuse is common across antisocial populations (Estévez & Emler, 2011). However, when we take the second-order factors of substance abuse into account, we do see a difference in the variables that measure more severe drug use (Drug use and Marijuana problems) between the *generic offenders* and the *impulsive-antisocial traits offenders* and the *psychopathic traits offenders.* This is in line with previous studies that suggested that the most antisocial and violent offenders also have the most severe drug problems (DeLisi, Vaughn, Salas-Wright, & Jennings, 2015; Sacks et al., 2009).

Finally, with regard to offending, the *non-antisocial psychopathic traits offenders, impulsive-antisocial traits offenders* and the *psychopathic traits offenders* significantly higher rates for both violent and non-violent crime than the *generic offenders*, which is consistent with previous findings (Hare & Neumann, 2008; Porter, Birt, & Boer, 2001). The *impulsive-antisocial traits offenders* scored significantly higher on violent crimes compared to the other profiles, possibly reflecting the poor behavioural control believed to typify these individuals (DeLisi, Tostlebe, Burgason, Heirigs, & Vaughn, 2016; Morgan & Lilienfeld, 2000).

Taken together, our findings showed that the *generic offenders* were characterized by relatively lower levels of psychopathic traits and lower scores on the external correlates examined (see Figure 2.2). *Impulsive-antisocial traits offenders* were typified by aggressive-, impulsive- and irresponsible behaviour, and they committed the highest amount of violent crimes compared to the other profiles. The *non-antisocial psychopathic traits offenders* were characterized by relatively high scores on the interpersonal, affective and lifestyle facets of psychopathy, but also on measures of anxiety, negative affect and externalizing. In contrast, they scored low on antisocial behaviour and violent crimes. The *psychopathic traits offenders* were characterized by high scores on all facets of psychopathy as well as on anxiety, negative affect, motivational tendencies and externalizing. The characteristics of the latter group were similar to that of the *non-antisocial psychopathic traits offenders*, but the *psychopathic traits offenders* showed significantly higher rates of criminal and aggressive behaviours.

An important limitation of the present study is that our analyses only included self-reported behavioural tendencies, which were used to infer psychological and personality traits. On one hand, this increases our insight on the classification of antisocial behaviour based on personality traits and observable behaviours, while on the other hand it precludes the identification of the biological and cognitive disturbances that underlie antisocial and psychopathic behaviours that characterize each group. This issue has been suggested to be one reason why therapeutic interventions are often ineffective in antisocial populations (Brazil et al., 2016). Another issue is that although the identified profiles were based on psychopathy scores, we cannot draw conclusions that are specific to the subtyping of psychopathy. One approach to gain more insight into subtypes within the population of offenders with psychopathy would be to include the relevant external correlates into separate latent profile analyses along with PCL-R scores (Kimonis, Fanti, Isoma, & Donoghue, 2013; Kimonis, Goulter, Hawes, Wilbur, & Groer, 2017). This would provide an opportunity to examine if homogenous subgroups of psychopathy emerge based on, e.g., negative affect or BIS/BAS.

With regard to the methodology, Tein and colleagues (2013) evaluated a selection of statistical parameters to determine the optimal number of classes and examined the power related to interclass distance between the classes. The present study evaluated a combination of statistical parameters, namely the BIC, LMR and entropy value, to determine the optimal number of classes. According to (Tein et al., 2013), the interclass distance (or Cohen's d) should not be lower than 0.08. Except for three comparisons, the Cohen's d was > 0.8. The three effect sizes that were <0.8 were SRP lifestyle factor (Cohen's d=-.25) for *impulsive-antisocial traits offenders* vs. *non-antisocial psychopathic traits offenders*, SRP lifestyle (Cohen's d=.68) for *non-antisocial psychopathic traits offenders* vs. *psychopathic traits offenders*, and SRP antisocial (Cohen's d=.56) for *impulsive-antisocial traits offenders* vs. *psychopathic traits offenders*.

In summary, the present study provides support for the presence of four distinct antisocial profiles based on self-report psychopathy scores in a male offender sample. The results are in line with previous findings that were based on clinical measures of psychopathy (Hare, 2016; Mokros et al., 2015). Furthermore, the present study provides relatively extensive and multifaceted characterizations of each profile. Since this is the first subtyping study using the SRP-SF, additional studies are required to further support replicability of the findings. In the future, profiles that are not only well-characterized in terms of personality correlates, but also incorporate biological and cognitive dimensions, could provide a more complete view of the individual (Brazil et al., 2016). This will aid diagnosis and foster the development of personalized treatment programs for individuals showing severe antisocial behaviour (Baskin-Sommers et al., 2015).

2.5 Supplementary Materials

Table S2.1 Bayes factors for the pairwise comparisons between profiles of the SRP facet scores and external variables.

	GE vs. IA		GE vs. NA		GE vs. PS	
	BF	Cohen's d	BF	Cohen's d	BF	Cohen's d
SRP Interpersonal	3.17E+15	-1.14	7.29E+48	-1.96	1.83E+74	-2.84
SRP Affective	9.95E+10	-0.96	1.22E+49	-1.97	5.97E+78	-2.97
SRP Lifestyle	1.12E+12	-1.00	4.32E+23	-1.27	7.27E+41	-1.91
SRP Antisocial	5.44E+76	-3.08	1.17E+14	-0.97	1.46E+93	-3.41
WAIS IQ	1.60E-1	0.07	1.38E-1	-0.05	1.40E-1	-0.04
APD	5.26E+20	-1.38	5.08E+7	-0.73	6.35E+33	-1.71
Violent crimes	1.63E+5	-0.7	1.23E-1	0.01	4.63E+1	-0.42
Non-violent crimes	4.88E+1	-0.45	8.81E+1	-0.42	3.51E+2	-0.49
Welsh-anxiety	2.79E+3	-0.58	9.25E+3	-0.55	1.46E+9	-0.85
MPQ-positive	1.60E-1	-0.07	1.47E-1	-0.07	3.00E-1	-0.16
MPQ-negative	1.19E+7	-0.81	5.53E+18	-1.14	2.43E+36	-1.8
BIS	1.40E-1	0.02	1.77E+0	0.27	7.64E+0	0.35
BAS	9.40E-1	-0.26	1.85E+5	-0.62	5.52E+11	-0.97

	IA vs. NA		IA vs. PS		NP vs. PS	
	BF	Cohen's d	BF	Cohen's d	BF	Cohen's d
SRP Interpersonal	1.06E+6	-0.84	4.72E+19	1.64	7.66E+5	0.8
SRP Affective	1.84E+9	-1.03	1.61E+27	2.02	1.91E+8	0.94
SRP Lifestyle	6.43E-1	-0.25	2.21E+6	0.9	1.52E+4	0.68
SRP Antisocial	1.73E+35	2.26	1.36E+2	0.56	8.90E+43	2.53
WAIS IQ	2.19E-1	-0.12	2.10E-1	0.11	1.55E-1	-0.01
APD	1.22E+2	0.55	1.43E+0	0.32	4.19E+6	0.85
Violent crimes	3.81E+2	0.59	1.50E+0	-0.33	1.17E+1	0.42
Non-violent crimes	1.77E-1	0.08	1.65E-1	0.03	2.10E-1	0.11
Welsh-anxiety	1.64E-1	0.05	5.69E-1	0.25	1.55E+0	0.31
MPQ-positive	1.56E-1	0	1.94E-1	0.09	1.77E-1	0.08
MPQ-negative	2.22E+0	-0.34	2.08E+7	0.99	1.45E+3	0.62
BIS	5.10E-1	0.23	1.16E+0	-0.31	1.72E-1	-0.07
BAS	2.27E+0	-0.35	1.18E+3	0.68	2.38E+0	0.34

Notes. GE = generic offenders, IA = impulsive-antisocial traits offenders, NA=non-antisocial psychopathic traits offenders, PS = psychopathic traits offenders. Bayes Factors are indicated as BF>10 (very strong); 3<BF<10 (strong) and BF<3 (weak).

Table S2.2 Bayes factors for the pairwise comparisons between profiles of the externalizing measures.

	GE vs. IA		GE vs. NA		GE vs. PS	
	BF	Cohen's d	BF	Cohen's d	BF	Cohen's d
ESI-Disinhibition	1.13E+11	-1.24	6.44E+7	-0.95	1.32E+1	-0.46
Irresponsibility	8.23E+4	-0.84	5.29E+2	-0.60	5.35E+4	-0.79
Impatient urgency	4.36E+3	-0.73	3.26E+5	-0.81	1.25E+12	-1.08
Problematic impulsivity	2.07E+9	-1.14	4.74E+4	-0.75	3.68E+12	-1.27
Fraud	2.27E+10	-1.20	1.49E+8	-0.97	9.19E+16	-1.50
Boredom proneness	2.10E+0	-0.36	1.48E+0	-0.31	4.59E+3	-0.71
ESI-Callous Aggression	4.14E+12	-1.33	3.39E+14	-1.31	8.51E+16	-1.50
Low empathy	1.68E+0	-0.35	1.08E+7	-0.90	3.39E+22	-1.77
Relational aggression	2.89E+4	-0.80	2.45E+8	-0.99	3.00E+19	-1.62
Excitement seeking	3.20E+0	-0.39	7.33E+2	-0.61	9.12E+12	-1.13
Destructive aggression	4.37E+6	-0.96	9.00E+2	-0.62	2.53E+18	-1.42
Physical aggression	8.51E+15	-1.52	3.97E+7	-0.94	1.90E+27	-1.99
Honesty	6.33E+1	-0.56	1.01E+2	-0.53	3.39E+6	-0.92
Rebelliousness	4.33E+9	-1.16	2.92E+8	-0.99	1.28E+25	-1.89
ESI-Substance abuse	1.26E+1	-0.47	4.13E-1	-0.21	2.38E+37	-2.50
Alcohol problems	1.08E+0	-0.31	2.43E-1	-0.14	1.80E+0	-0.34
Drug use	8.11E+3	-0.76	7.14E+0	-0.41	3.71E+3	-0.70
Marijuana problems	8.41E+1	-0.57	2.47E-1	-0.14	1.12E+1	-0.45
Marijuana use	8.13E+0	-0.45	4.91E-1	-0.22	1.39E+0	-0.32

	IA vs. NA		IA vs. PS		NA vs. PS	
	BF	Cohen's d	BF	Cohen's d	BF	Cohen's d
ESI-Disinhibition	7.64E-1	0.31	7.32E-1	0.31	4.13E+1	0.60
Irresponsibility	4.16E-1	0.23	1.95E-1	-0.02	3.50E-1	0.20
Impatient urgency	1.96E-1	-0.05	9.70E-1	0.35	9.16E-1	0.32
Problematic impulsivity	2.71E+0	0.43	3.04E-1	0.18	3.16E+1	0.58
Fraud	6.35E-1	0.29	3.06E-1	0.18	1.28E+1	0.53
Boredom proneness	2.20E-1	0.06	9.87E-1	0.35	2.75E+0	0.42
ESI-Callous Aggression	2.00E-1	0.06	8.15E+4	1.02	1.49E+7	1.14
Low empathy	4.57E+0	-0.47	2.71E+6	-1.16	4.23E+2	0.72
Relational aggression	2.33E-1	-0.12	1.62E+2	0.72	1.48E+2	0.67
Excitement seeking	3.06E-1	-0.18	4.77E+1	0.65	6.53E+0	0.49
Destructive aggression	1.98E-1	0.34	2.28E+0	0.43	1.30E+3	0.78
Physical aggression	1.48E+1	0.55	4.94E+0	0.49	6.40E+5	1.03
Honesty	1.91E-1	0.03	7.48E-1	0.32	1.20E+0	0.35
Rebelliousness	3.07E-1	0.18	3.58E+1	0.63	1.04E+4	0.89
ESI-Substance abuse	5.60E-1	0.27	1.95E-1	-0.01	5.14E-1	0.26
Alcohol problems	9.84E-1	0.17	1.96E-1	0.03	3.44E-1	0.20
Drug use	1.33E+0	0.37	1.98E-1	-0.04	8.06E-1	0.31
Marijuana problems	2.36E+0	0.42	2.52E-1	-0.14	7.99E-1	0.31
Marijuana use	7.43E-1	0.31	4.09E-1	-0.23	2.27E-1	0.12

Notes. GE = generic offenders. IA = impulsive-antisocial traits offenders. NA=non-antisocial psychopathic traits offenders. PS = psychopathic traits offenders. Bayes Factors are indicated as BF>10 (very strong); 3<BF<10 (strong) and BF<3 (weak).

3

Psychopathic traits influence threat avoidance in a community sample independent of testosterone

This chapter is based on:
Driessen, J. M. A., Brazil, I. A., Dorta Lorenzo, E., Herwaarden, A. E., Olthaar, A. J., Potamianou, T. I., & Glennon, J. C. (in press). Psychopathic traits influence threat avoidance in a community sample independent of testosterone. *Personality Disorders: Theory, Research, and Treatment.*

Abstract

Psychopathy is a personality construct that encompasses a constellation of traits reflecting emotional dysfunction and antisocial behavior. Individuals with elevated levels of psychopathic traits have shown abnormal affective processing. Studies with psychopathic offenders suggested that this is a result of altered automatic social approach-avoidance tendencies. The goal of the current study was to increase the insight into the underlying mechanism of affective processes in community-dwelling individuals with a high level of psychopathic traits by studying approach and avoidance behavior in an experimental setting. Eighty-seven healthy individuals performed a computerized affective approach-avoidance task in which they pushed or pulled emotional faces using a joystick. The results showed that high levels of psychopathic traits corresponded with diminished threat avoidance to angry faces, as was found previously in psychopathic offenders. Furthermore, given its link with aggression and threat approach, testosterone was measured in order to investigate a possible mediatory role. Although testosterone was positively associated with psychopathic traits, endogenous testosterone did not mediate the effect of psychopathic traits on threat avoidance. We propose that an increased understanding of the interplay between different neuroendocrine mechanisms could lead to a better insight into the underlying mechanism of abnormal threat avoidance in individuals with high levels of psychopathic traits.

3.1 Introduction

Psychopathy is a disorder that is generally marked by emotional dysfunction and antisocial behavior (Hare, 1991; Lynam, Caspi, Moffitt, Loeber, & Stouthamer-Loeber, 2007). It has been demonstrated that psychopathy is a dimensional construct and psychopathic traits are distributed normally in the general population (Gao & Raine, 2010; Levenson et al., 1995). Previous research emphasizes abnormal affective processing in individuals with elevated levels of psychopathy (Gordon et al., 2004; Hare, 1999; Kiehl et al., 2001). Such individuals fail to fully experience or appreciate the affective significance of emotional stimuli differently in comparison to individuals with low levels of psychopathy (Christianson et al., 1996; Hastings, Tangney, & Stuewig, 2008; Kiehl, Hare, McDonald, & Brink, 1999; Kiehl et al., 2001; Patrick, Cuthbert, & Lang, 1994; Williamson, Harpur, & Hare, 1991). Neuroimaging studies demonstrate that processing of affective stimuli is associated with diminished activity in limbic structures and over-activation in the frontal cortex in psychopathic offenders (Kiehl et al., 2001; Müller et al., 2003). It has been suggested that abnormal processing of affective information may be the underlying source of aggression, predatory violence and deviant threat responses associated with psychopathy (Reidy, Shelley-Tremblay, & Lilienfeld, 2011; Skeem & Mulvey, 2001; von Borries, Brazil, Bulten, Buitelaar, Verkes, & de Bruijn, 2012). One way to study affective processing is by investigating approach and avoidance behavior under different conditions.

Approach and avoidance reactions are basic responses to emotional stimuli and underlie every complex emotional response (Lang, Bradley, & Cuthbert, 1997). The natural tendency is to approach pleasant, positive stimuli and to avoid unpleasant, negative stimuli (Lang & Bradley, 2010; Puca, Rinkenauer, & Breidenstein, 2006). Studies have shown that approach is associated with pulling objects closer while avoidance is associated with pushing objects away from oneself (Chen & Bargh, 1999; Solarz, 1960). People appear to respond faster and make less errors when they have to make an affect-congruent response (approach a positive stimulus, avoid a negative stimulus) in comparison to an affect-incongruent response (vice versa) (Roelofs, 2017; Roelofs, Minelli, Mars, Van Peer, & Toni, 2008; Stins et al., 2011). It was suggested that an affect-congruent condition requires responses that are consistent with the automatic tendencies (approach-positive, avoid-negative), while these automatic tendencies need to be controlled during the affect-incongruent condition, in order to apply the counterintuitive action of approaching negative and avoiding positive stimuli (Chen & Bargh, 1999; Roelofs et al., 2008). Prefrontal control and motivational networks have found to be involved in approach and avoidance reactions (Ernst & Fudge, 2009; Hariri, Tessitore, Mattay, Fera, & Weinberger, 2002). More specifically, the prefrontal cortex (PFC), and particularly the anterior cingulate cortex is suggested to be crucial for overriding automatic action tendencies, supposedly via modulation of the amygdala and periaqueductal grey matter (Roelofs, 2017; Volman, Toni, Verhagen, & Roelofs, 2011). Interestingly, psychopathic offenders showed a lack of threat avoidance tendencies in response to angry faces. This effect was shown to be inversely related to the level of instrumental aggression (von Borries et al., 2012). Moreover, offenders with psychopathic traits showed reduced PFC activation as well as PFC-amygdala connectivity during affect-incongruent behavior (Volman et al., 2016a). These findings indicate that altered automatic action tendencies might underlie disturbed emotional processing (Volman et al., 2016a; von Borries et al., 2012). Abnormalities in affective processing were also found in non-offenders with elevated levels of psychopathic traits. For instance, men reporting high levels of psychopathic traits showed atypical startle reflexes when viewing aversive stimuli (Justus & Finn, 2007). Moreover, individuals with a high level of psychopathic traits showed altered frontal cortex and amygdala responses to emotional expressions in comparison with medium and low tendencies (Gordon et al., 2004). It is yet unclear whether the same mechanisms underlie altered affective behavior in offenders and non-offenders with psychopathic traits. So far, studies employing a dimensional approach to psychopathy have been mainly focused on perceptual processing of emotional expressions. Thus, it is not clear whether the affective problems associated with psychopathic traits in the general population are manifested at the basic level of automatic approach-avoidance tendencies.

Although the exact mechanism underlying approach-avoidance behavior remains unclear, previous studies have emphasized a role for the neuroendocrine system (Kaldewaij, Koch, Volman, Toni, & Roelofs, 2016). Testosterone is a hormone that is strongly associated with aggressive and approach-related behavior (Book, Starzyk, & Quinsey, 2001; Lombardo et al., 2012). For instance, testosterone administration studies highlighted the testosterone-initiated bias toward the approach of social threat, that may underlie mechanisms of social dominance and aggression (Enter, Spinhoven, & Roelofs, 2014; Enter, Spinhoven, & Roelofs, 2016; Radke et al., 2015; Volman et al., 2016a). Previous studies have shown that endogenous testosterone levels were inversely related to PFC activity and PFC-amygdala connectivity during affect-incongruent trials (Volman et al., 2011). This was further supported by findings that indicated a negative association related to reduced PFC and amygdala activation, as well as decreased connectivity between the two, in both cognitively unimpaired (Mehta & Beer, 2010; Peper, van den Heuvel, Mandl, Pol, & van Honk, 2011; Waller, Gard, Shaw, Forbes, Neumann, & Hyde, 2019) and psychopathic individuals (Volman et al., 2016a). Importantly, reduced threat sensitivity, increased dominance, and excessive aggression, which can be triggered by testosterone, are key features of psychopathy. Taken together,

it seems that a testosterone-initiated increase in approach behavior towards threatening stimuli is comparable to the increase of approach behavior towards threats that has been reported in psychopathic offenders. Notably, it remains unclear whether endogenous testosterone serves a mediatory role in these altered automatic action tendencies.

In the present study, we aimed to investigate the link between psychopathic traits and approach-avoidance behavior in the general population. Participants were included in the study based on their level of psychopathic traits, measured with the Self-Report Psychopathy scale (Gordts, Uzieblo, Neumann, Van den Bussche, & Rossi, 2017) and performed an affective approach-avoidance task (AAT). A substantial amount of literature has demonstrated the potential of the approach-avoidance paradigm in capturing the automatic behavioral tendencies towards emotional expressions in a variety of samples, such as healthy individuals (Chen & Bargh, 1999; Enter et al., 2014; Marsh, Ambady, & Kleck, 2005; Radke et al., 2016; Radke et al., 2015; Roelofs et al., 2008; Rotteveel & Phaf, 2004; Seidel, Habel, Kirschner, Gur, & Derntl, 2010; von Borries et al., 2012), individuals with high anxiety (Heuer, Rinck, & Becker, 2007; Lange, Keijsers, Becker, & Rinck, 2008; Lange et al., 2010; Rinck & Becker, 2007; Roelofs et al., 2010; Struijs et al., 2017), and phobia (Klein, Becker, & Rinck, 2011; Rinck & Becker, 2007). In line with previous research, we expected to find a negative link between psychopathic traits and threat avoidance (von Borries et al., 2012). In addition, we tested how the different facets of the SRP were related to approach-avoidance behavior. The SRP distinguishes four facets; interpersonal (arrogant and deceitful interpersonal style), affective (deficient affective experience), lifestyle (impulsive-irresponsible behavioral style), and antisocial (antisocial traits)(Hare & Neumann, 2005). We expected to find an effect on threat avoidance for all the facets, as interpersonal and affective traits represent deficient emotional experience and social dominance, while lifestyle and antisocial traits are associated with impulsivity and lack of fear. All of these aspects were expected to play a role in social interaction and affective processing. Furthermore, we measured salivary testosterone in order to investigate a possible mediation effect of baseline testosterone on approach-avoidance behavior as a function of psychopathic traits.

3.2 Methods

3.2.1 Participants

Participants between 18 and 35 years were recruited (*N*=87, male=36, M_{age}=24.09, SD_{age}=2.68) via local advertisement and an online research participation system (SONA systems). They were native Dutch speakers and exclusion criteria for the study were self-reported epilepsy, brain surgery, claustrophobia, history of other neurological conditions or psychiatric disease, use of psychoactive medication or substances, and pregnancy. To evaluate their psychopathic traits, participants were assessed with the short form of the self-report psychopathy scale (SRP-SF; Dutch version: (Gordts et al., 2017). The SRP-SF is a self-report questionnaire including 28 items that need to be rated on a 5-point Likert scale. A substantial amount of literature supports the validity and reliability of the SRP-SF in a community sample (Gordts et al., 2017; Hare & Neumann, 2008; Lilienfeld, Fowler, & Patrick, 2006; Neumann & Dustin Pardini, 2014). The total score as well as the scores on the individual facets interpersonal (Int), affective (Aff), lifestyle (Lif), and antisocial (Ant) could be derived.

Participants were included based on their SRP total score, in order to obtain a reliable distribution of psychopathic traits in our test sample. A questionnaire database including the SRP (selection sample with *N*=1519, 309 males) was used to define a typical distribution in the general population and the total scores were divided in quartiles. Within the test sample, 25% of the participants belonged to the top and bottom quartiles, while 50% of the participants belonged to the two middle quartiles. The top and bottom quartiles were oversampled in order to enhance the presence of extreme scores on both sides of the distribution (Bernat, Nelson, Steele, Gehring, & Patrick, 2011; Brazil, Hunt, et al., 2013; Gong, Brazil, Chang, & Sanfey, 2019). Internal consistency in both the test sample and the selection sample was good (test sample: Int α = .82; Aff α = .69; Lif α = .78; Ant α = .83; selection sample: Int α = .81; Aff α = .75; Lif α = .78; Ant α = .82). The test sample showed a similar distribution compared to the selection sample (Figure S3.1). Participants provided written informed consent prior to the onset of the study, and received monetary compensation at the end of the study. Procedures were approved by the Ethics Committee Social Sciences of the University of Nijmegen.

3.2.2 Task

The Approach- Avoidance Task is a well-validated computerized task that is designed to measure the automatic approach-avoidance behavior (e.g., Enter et al., 2014; Marsh et al., 2005). The stimuli used in this paradigm consisted of 84 gray scale photographs of male and female actors using angry or happy facial expressions with direct gaze (Radboud Faces Database). All pictures were resized to a size of 1024×681 pixels and presented against a black background. Variation in physical parameters unrelated to the emotional expression was minimized by excluding the hair and other non-facial aspects and by matching the pictures in sizes, brightness and contrast. In 16 blocks of 12 trials each, stimuli with different emotional valence (happy or angry) were presented on a computer screen. Stimuli were presented for 100ms, preceded by a fixation cross in the center of the screen.

A joystick (Logitech Attack 300) was located between the computer screen and the participant. The participant was informed about two alternating sets of instructions that changed over blocks: 1) pull-happy, push-angry (congruent condition) and 2) push-happy, pull-angry (incongruent condition). The participant was instructed to respond as accurately and quickly as possible. The maximum response time was 2s. The actual experiment was preceded by a practice session of four blocks with eight stimuli each. In order to test the internal consistency of the behavioral responses, Cronbach alpha's were calculated over the minimum amount of correct trials of each condition (k=30 trials). The internal consistency was high in all conditions (pull-happy: α = .95; push-happy: α = .92; push-happy: α = .93; pull-angry: α = .94).

3.2.3 Testosterone measure

At the end of the study, we asked participants to fill one collection kit (Oragene saliva (RE-100), DNA Genotek, Ltd.) with saliva, which was stored at -80°. Participants were asked to abstain from consuming food and drinks other than water for at least 2.5 hours before the saliva collection in order to minimize the impact of variables that could influence testosterone levels. Saliva was collected at the end of the test session and, where possible, the test sessions were performed within the same time window.

Testosterone was analyzed by liquid chromatography-tandem mass spectrometry (LCMS/MS) after solid-phase extraction. Briefly, 500 µL saliva was extracted (Oasis MCX 1cc, Waters) after addition of internal standard [13C3] c-testosterone (Isoscience, King of Prussia, PA). The eluate was injected into an Agilent Technologies 1290 Infinity VL UHPLC-system equipped with a BEH C18 (1.7µm 2.1 X 50mm) analytical column (Waters Corp.) coupled to an Agilent 6490 tandem mass spectrometer (Agilent Technologies, Santa Clara, CA). The method was linear assessed by CLSI EP6 protocol (Tholen et al., 2003). Testosterone was quantified with an 9-point calibration curve (Steraloids) and the method has a total CV of 5.5% at 0.09 nmol/L and 4.6% at 0.49 nmol/L.

3.2.4 Behavioral analyses

Reaction times and amount of errors for each condition were extracted from the data while using MATLAB (Mathworks). Reaction times (RT) shorter than 150ms and longer than 1500ms were considered outliers and removed (von Borries et al. 2012). In order to investigate the expected congruency effect, two repeated measures ANOVA were performed with Movement (push, pull) and Emotion (happy, angry) as Within-Subject variables for reaction time and amount of errors. The AAT effect scores were defined as the difference in RT between push and pull reactions for happy and angry stimuli. We created latent variables for the AAT effect scores in Mplus 7 (Muthén & Muthén, 1998–2015) in order to reduce the measurement error of the observed variables. The average effect scores for happy and angry affect were loaded onto the corresponding latent variables. We exported the factor loadings for the two latent variables of each participant to MATLAB and used them for further analyses. Subsequently, we analyzed the correlations of the latent AAT effect scores with the psychopathy subscale scores and the total psychopathy score. Non-parametric tests were used as the data were not normally distributed.

To test for a possible mediation effect of testosterone on the relation between SRP total score and latent angry effect score, we performed a bootstrapped mediation analysis in Mplus 7 using the maximum-likelihood method and 5000 random samplings of the data, as recommended by Hayes (2009). Bootstrapping analysis is more powerful than testing the significance of indirect effects through the conservatively biased Sobel test (Sobel, 1982). Significance levels (alpha) were set to 0.05 for all analyses.

3.3 Results

Mean RT of correct trials and error rates are presented in Table 3.1. An interaction effect for movement and emotion on reaction time (F(1,86)=84.24, p<.001, η_p^2=0.50, *90%CI*=[0.37, 0.59]) and error rate (F(1,86)=22.62, p<.001, η_p^2=0.21, *90%CI*=[0.09, 0.32]) indicated that participants responded significantly slower (M=665.33, SE=12.31) and made more errors in response to incongruent trials (M=7.91, SE=0.65) in comparison to congruent trials (RT: M=641.48, SE=11.88; errors: M=5.13, SE=0.57) (Figure 3.1). Furthermore, there was a main effect on reaction time for Movement (F(1,86)=33.08, p<.001, η_p^2=0.28, *90%CI*=[0.15, 0.39]) and errors (F(1,86)=21.92, p<.001, η_p^2=0.20, *90%CI*=[0.09, 0.31]), indicating that participants made faster

Table 3.1 Reaction times and amount of errors for each condition.

		Approach	Avoidance
		M(SE)	M(SE)
Reaction time	Happy	623.20 (12.69)	679.57 (11.87)
	Angry	651.08 (13.14)	659.76 (11.45)
Errors	Happy	2.25 (0.28)	4.61 (0.39)
	Angry	3.30 (0. 32)	2.87 (0.34)

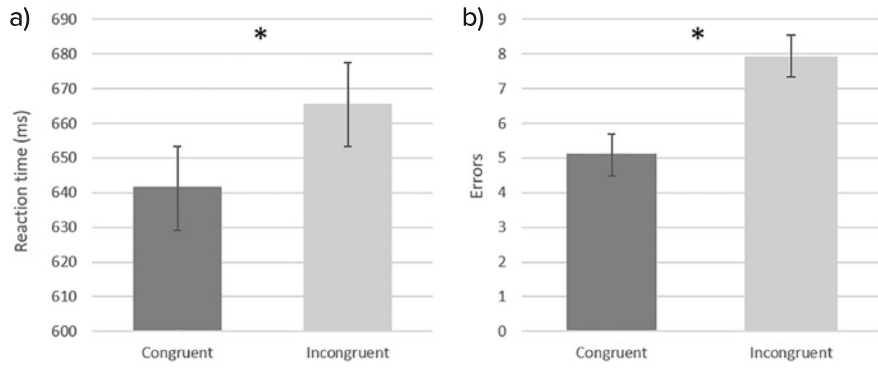


Figure 3.1 Congruency effect for **a)** reaction times and **b)** errors. *p<.05, error bars represent standard error of the mean.

responses and less errors when making a pull response (RT: M=637.14, SE=11.36, errors: M=5.55, SE=0.50) compared to a push response (RT: M=669.66, SE=12.70, errors: M=7.48, SE=0.63). There was no significant main effect of Emotion (F(1,86)=1.99, p=.169, η_p^2=0.02, *90%CI*=[0, 0.09]) and errors (F(1,86)=3.69, p=.058, η_p^2=0.04, *90%CI*=[0, 0.13]).

The AAT effect scores for angry stimuli were positively correlated with SRP total scores (Spearman's rho (ρ)=.33, *90%CI*=[0.16, 0.48], p=.002,) (Figure 3.2). The subscales interpersonal (ρ=.42, *90%CI*=[0.26, 0.56], p<.001), lifestyle (ρ=.32,

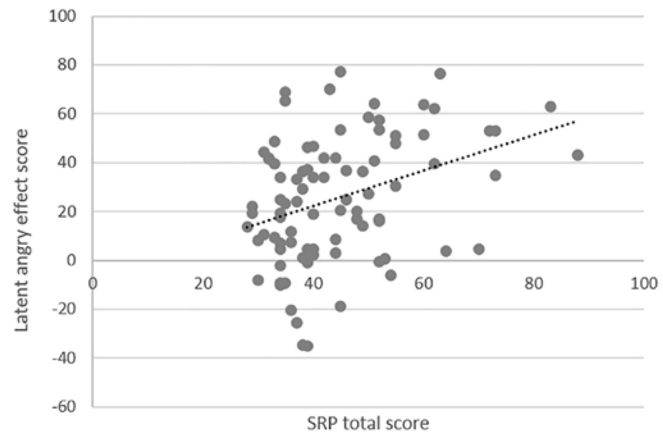


Figure 3.2 Positive correlation SRP total score and latent angry effect score for angry faces.

90%CI=[0.15, 0.47], *p*=.003) and antisocial (ρ=.32, *90%CI*=[0.15, 0.47], *p*=.003) also correlated positively with the AAT effect scores. There was no correlation with the affective subscale (ρ=.11, *90%CI*=[-0.07, 0.28], *p*=.315). The AAT effect scores for happy stimuli were not significantly correlated with the SRP total scores or the subscale scores (Table 3.2).

Table 3.2 Spearman pairwise correlations between SRP scores and latent effect scores.

	Happy			Angry		
	rho	p	90% CI	rho	p	90% CI
SRP Total	.02	.839	[-.16, .19]	.32	.002**	[.16, .48]
SRP Interpersonal	-.04	.730	[-.21, .14]	.42	<.001***	[.26, .56]
SRP Affective	.05	.653	[-.13, .23]	.11	.315	[-.07, .28]
SRP Lifestyle	.08	.445	[-.10, .26]	.32	.003**	[.15, .47]
SRP Antisocial	.10	.368	[-.08, .27]	.32	.003**	[.15, .47]

Notes. Asterisks indicates significant correlations (***p*<.01, ****p*<.001).

In order to investigate whether testosterone mediates the relation between SRP total score and the latent angry effect scores, a mediation model was tested (Figure 3.3). The results indicated that SRP total score predicted the latent angry affect scores (β=0.65, *90%CI*=[0.35, 1.00], *SE*=0.20, *p*=.001, β*=0.32), and testosterone (β=0.003, *90%CI*=[0.002, 0.004], *SE*=0.001, *p*<.001, β*=0.36). However, testosterone was not significantly related to the latent angry effect scores (β=29.21, *90%CI*=[-23.78, 91.70], *SE*=31.71, *p*=.357, β*=0.10). While the direct effect of SRP total score on latent angry effect score was significant (β=0.65, *90%CI*=[0.35, 0.99], *p*=.001, β*=0.32), the indirect effect was not (β=0.08, *90%CI*=[-0.07, 0.22], *p*=.373, β*=0.04). These findings indicate that testosterone did not mediate the relation between SRP total score and the latent angry effect score. Analyses with the SRP facets also showed non-significant findings which suggested testosterone did not mediate the relation between the SRP facets and the latent angry effect score.

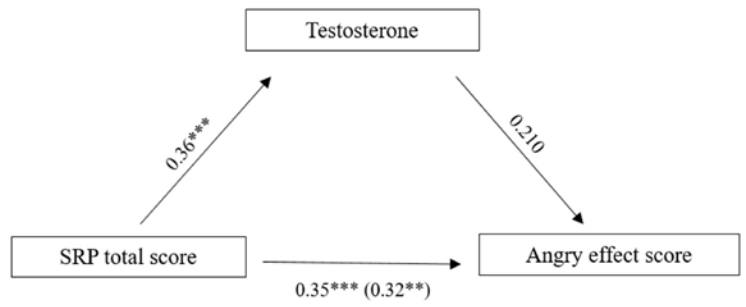


Figure 3.3 Standardized regression coefficients and level of significance (**p*<.05, ***p*<.01, ****p*<.001) for the relationship between SRP total score, latent angry effect score and testosterone. Along the lower path, the number outside parentheses is the total effect, and the number inside parentheses is the direct effect.

3.4 Discussion

The aim of the present study was to investigate the effect of psychopathic traits on approach-avoidance behavior in a community sample. In addition, we investigated whether endogenous testosterone levels associated with psychopathy could be related to alterations in approach-avoidance behavior. Our findings indicated that individuals with a high tendency towards psychopathy lack avoidance of angry faces. We did not find evidence for a mediatory role of testosterone.

In most people, confrontation with a threatening stimulus, e.g. an angry face, elicits personal distress and initiates an avoidance response (Lang et al., 1997). Lack of threat avoidance in response to angry faces was reported in psychopathic offenders (von Borries et al., 2012), which is in line with models suggesting low threat reactivity in psychopathic individuals (Hoppenbrouwers, Bulten, & Brazil, 2016). Current findings indicate that a lack of automatic avoidance tendencies towards anger is also present in non-incarcerated individuals with elevated levels of psychopathy. More specifically, these individuals were faster in approach responses towards angry facial expressions in comparison with avoidance responses. Thus, while previous studies suggested reduced reactivity towards sad and fearful facial expressions, reactivity towards angry expressions seemed to be elevated. This could be explained by the phenotypic features of psychopathy, namely interpersonal dominance, reduced threat sensitivity, and impulsivity. While antisocial behavior is more prevalent in (psychopathic) offenders, interpersonal and affective psychopathy traits are prominent in both offenders and non-offenders with psychopathic tendencies. Therefore, it is not surprising that our findings indicate that the same mechanisms underlie altered affective behavior in offenders and non-offenders with psychopathic traits. Where interpersonal dominance involves the urge to be in charge and control over the environment (e.g. influence or direct other people) (Pratto, Sidanius, Stallworth, & Malle, 1994), diminished threat sensitivity and impulsivity are associated with the reduced control over emotional reactions.

We noted that mean RTs for congruent and incongruent trials in the happy affect condition were not very different. As expected from these results, post-hoc tests revealed that the congruency effect was mainly driven by the happy affect condition. We speculate that the effect of psychopathic traits on threat avoidance influenced the congruency effect in our sample. That is, our results indicate that psychopathic traits are negatively linked to automatic threat avoidance, and are in line with previous findings (von Borries et al., 2012). This suggests that the congruency effect for angry faces was impaired in these individuals. The presence of a relatively high number of individuals with psychopathic traits in our sample could be an explanation for the absence of a significant congruency effect for the angry affect condition. Besides, note that we did not find a link between the affective subscale of the SRP and threat avoidance. Affective traits reflect the experience of emotions, while the AAT reflects the processing of emotional stimuli and the control over these processes. The PFC and the amygdala are key areas that are involved in emotional control (Etkin, Büchel, & Gross, 2015), and psychopathic offenders have shown reduced PFC activation and reduced amygdala-PFC functional connectivity during executive control (Volman et al., 2016b). Therefore, follow-up studies investigating PFC and amygdala activity during emotional control in non-offenders with psychopathic traits are needed.

Disturbed emotional processing as a result of an imbalance between subcortical and cortical areas, or of impaired communication between the two regions, was previously suggested in the triple balance model of emotion which distinguishes three imbalances of emotional processing in psychopaths (van Honk & Schutter, 2006). According to this model, increased testosterone levels could reduce the communication between subcortical and cortical areas, resulting in diminished control of motivational tendencies. However, we did not find that testosterone mediated the relation between psychopathic traits and approach-avoidance behavior. So, our results do not support the triple balance hypothesis. In line with previous studies (Glenn, Raine, Schug, Gao, & Granger, 2011; Welker, Lozoya, Campbell, Neumann, & Carré, 2014), we found a positive association between testosterone and psychopathy. One important note is that all studies that reported a positive link were conducted in community samples. It could be that community samples represent a broader distribution of testosterone levels, while testosterone levels in offenders are relatively high (Fang et al., 2009; Maras et al., 2003). As a

consequence, correlations remain undetected in offenders. Another important hormone included in the triple balance model is cortisol. It was proposed that the cortisol-testosterone ratio influences activity in the subcortical areas, such as the amygdala. In addition, while testosterone was suggested to reduce communication between cortical and subcortical areas, cortisol was believed to have the opposite effect. This prediction is in line with findings suggesting that cortisol may directly moderate the behavioral effects of testosterone (Dabbs, Jurkovic, & Frady, 1991; Popma et al., 2007). This might point towards a compensatory mechanism of cortisol on the effect of testosterone on approach-avoidance behavior in individuals with elevated levels of psychopathic traits. Therefore, it could be interesting to consider the cortisol-testosterone ratio in future studies. Finally, it should be noted that the fact that we did not find a mediation effect of testosterone on the relation between psychopathic traits and approach-avoidance behavior could be due to sample size limitations. Future studies should aim for larger samples in order to confirm our findings.

In conclusion, the current study is the first to investigate the effects of psychopathic traits and testosterone on approach-avoidance behavior in a general population sample. Our results indicate that individuals with high levels of psychopathic traits show a lack of automatic threat avoidance, as was found previously in psychopathic offenders. Although testosterone was positively related to psychopathic traits, it did not mediate the effect of psychopathic traits on threat avoidance. A better understanding of the interplay between different neuroendocrine mechanisms could lead to a better insight into the lack of threat avoidance in individuals with high levels of psychopathic traits.

3.5 Supplementary material

a)

b)

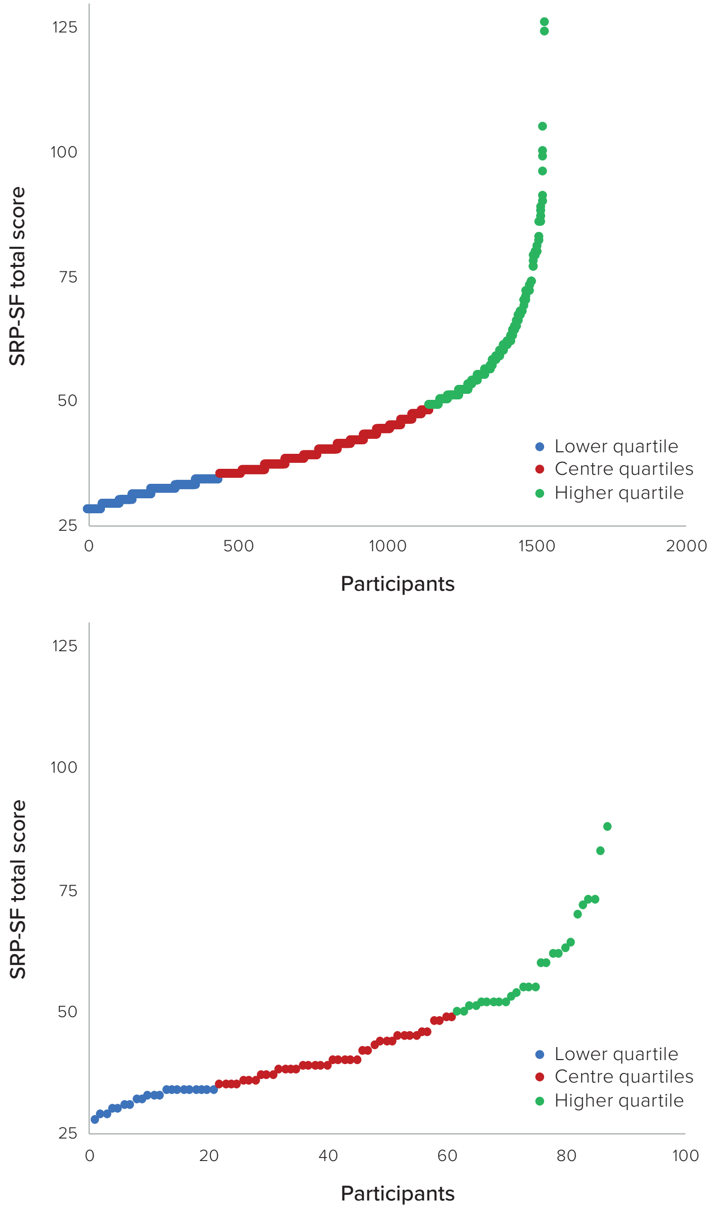


Figure S3.1. Distributions of SRP total scores in **a)** the larger community sample (*N*=1519) and **b)** the test sample (*N*=87).

4

Pain processing in a social context and the link with psychopathic personality traits

This chapter is based on:
Van Heck, C. H.*, Driessen, J.M.A.*, Amato, M., van den Berg, M. N., Bhandari, P., Bilbao-Broch, L., ... Jongsma, M. L. A. (2017). Pain processing in a social context and the link with psychopathic personality traits – an event-related potential study. *Frontiers in Behavioral Neuroscience, 11,* 180. **shared first authorship*

Abstract

Empathy describes the ability to understand another person's feelings. Psychopathy is a disorder that is characterized by a lack of empathy. Therefore, empathy and psychopathy are interesting traits to investigate with respect to experiencing and observing pain. The present study aimed to investigate pain empathy and pain sensitivity by measuring event-related potentials (ERPs) extracted from the ongoing EEG in an interactive setup. Each participant fulfilled subsequently the role of 'villain' and 'victim'. In addition, mode of control was modulated resulting in four different conditions; passive villain, active villain, active victim and passive victim. Response-, visual- and pain ERPs were compared between the four conditions. Furthermore, the role of psychopathic traits in these outcomes was investigated. Our findings suggested that people experience more conflict when hurting someone else than hurting themselves. Furthermore, our results indicated that self-controlled pain was experienced as more painful than uncontrolled pain. People that scored high on psychopathic traits seemed to process and experience pain differently than those low on psychopathic traits. According to the results of the current study, social context, attention and personality traits seem to modulate pain processing and the empathic response to pain in self and others. The within-subject experimental design described here provides an excellent approach to further unravel the influence of social context and personality traits on social cognition.

4.1 Introduction

4.1.1 Pain and empathy

From an evolutionary point of view, pain signals actual or potential injury or damage to bodily parts and is thereby a protective mechanism. Perceived pain severity can be greatly influenced by various factors, such as attention and expectancy (Melzack & Wall, 1967). Moreover, it has been determined that pain is perceived as less intense when it is self-controlled (Pellino & Ward, 1998; Salomons, Johnstone, Backonja, & Davidson, 2004).

Humans are naturally social individuals and experience discomfort while observing another person in pain. This phenomenon, termed as 'pain empathy', is a complex construct that describes the ability to understand another person's situation or feelings (Davis, 1980; Lietz et al., 2011) and is believed to be one of the requirements for successful participation in current society (Schneider & Ingram, 2005). Neuroimaging studies focussing on empathy received considerable effort in the past decade (Decety, 2010; Singer & Lamm, 2009). For instance, previous studies showed that ongoing information processing is affected differently when being exposed to pictures that show another person's pain than being exposed to neutral pictures (Avenanti, Bueti, Galati, & Aglioti, 2005; Bufalari, Aprile, Avenanti, Di Russo, & Aglioti, 2007). Evidence from neuroimaging research suggests that experienced pain and observed pain in others elicit similar activation patterns in brain areas involved in the processing of both affective (e.g. the anterior insula and the medial/anterior cingulate cortex (Decety, 2010)) and sensory (e.g. the primary somatosensory cortex and parietal operculum (Bufalari et al., 2007) information. These findings support the theory that describes a shared neural network for one's own and others' emotional and sensory experience.

Current models of pain empathy suggest that empathy-related processes are derived from both bottom-up features and top-down factors (Decety & Moriguchi, 2007). Zooming in on these top-down factors, social context seems to be an important modulator of pain perception in self and others (Decety, Michalska, Akitsuki, & Lahey, 2009; Singer et al., 2006). Several aspects of social context, such as relationships between individuals (Singer et al., 2006) and attitude towards others (Decety et al., 2009) have been studied previously. However, studying the lack of empathy with respect to pain might be even more salient.

4.1.2 Psychopathy; a pain- and empathy-related disorder

Psychopathy is a disorder that is linked to deviant pain processing and experience. Although the majority of studies have been focused on psychopathy in criminal offenders (Thompson, Ramos, & Willett, 2014), psychopathic personality traits are demonstrated to be normally distributed in the general population (Gao & Raine, 2010; Hare & Neumann, 2008; Levenson et al., 1995). Recent neuroimaging studies have suggested that an attenuated function in the amygdala and anterior insula underlies reduced empathy in individuals with high levels of psychopathic traits (Seara-Cardoso et al., 2015). Moreover, research revealed that people high in psychopathic traits show atypical neural activity in response to imagining others' pain (Decety, Chen, Harenski, & Kiehl, 2013; Seara-Cardoso et al., 2015). Besides characteristics of lack of empathy, psychopaths tend to experience pain differently compared to non-psychopaths. For instance, Marcoux and colleagues (2014) found a higher pain threshold in people with psychopathic tendencies.

4.1.3 Electrophysiology in pain research

Electrophysiological techniques, such as EEG, can discriminate event-related activity with a high temporal resolution and are therefore excellent methods to study if and when differences in neural signals related to certain events occur. Extracting such event-related activity from the ongoing electroencephalogram (EEG) allows researchers to study event-related potentials (ERPs). ERPs can be elicited by either actions, simple or complex stimuli, or events. This ERP technique allows us to directly study the neural responses associated with specific aspects of emotion and information processing.

4.1.3.1 Response-locked ERPs

A specific component of the response-locked ERP that is studied in empathy-related research is the error-related negativity (ERN), which is an event-related potential that is associated with an incorrect motor response (e.g. a button press). It starts shortly before the time of an incorrect response and peaks around 100ms thereafter (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). The ERN is generated within or near the dorsal anterior cingulate cortex (ACC; Dehaene, Posner, & Tucker, 1994). Electrophysiological evidence demonstrated an association between the ERN, as electrophysiological correlate of action monitoring, and empathy-related affective responding (Larson, Fair, Good, & Baldwin, 2010; Thoma & Bellebaum, 2011). According to different theories, the ERN reflects the error-detection process itself (Falkenstein et al., 1991), or an emotional response to the error (Bush, Luu, & Posner, 2000). Regarding the latter, research showed that an increased ERN has been associated with, for instance, concern over the outcome of an event (Gehring, Himle, & Nisenson, 2000; Gehring & Willoughby, 2002). In line, a diminished ERN has been associated with a lack of concern over the outcome of an event (Santesso & Segalowitz, 2009).

4.1.3.2 Visual ERPs

Visual stimuli result in a series of peaks in the EEG and thereby determine the visual ERP. Perhaps the most studied component with respect to a wide range of cognitive processes is the P300 component (or P3). This visual P3 component is modulated by cognitive processes such as expectancy, relevance, meaning and attention (Gray, Ambady, Lowenthal, & Deldin, 2004). Several studies found that viewing painful stimuli caused a larger visual P3 amplitude over the posterior parietal area compared to viewing neutral pictures (Fan & Han, 2008; Meng et al., 2012).

4.1.3.3 Pain ERPs

Previous literature on empathy is mostly based on studies in which participants are not exposed to actual pain or pain in others directly (Botvinick et al., 2005; Singer et al., 2004). A more realistic, though controversial method, would be to introduce real-life situations of pain experience. Such experimental setups are not very common. One famous example stems from the controversial Milgram experiment that studied obedience (Milgram, 1963). In the current study, we adapted this approach to investigate the processing of painful stimuli delivered to oneself or to another person in both an active and a passive condition. Electrophysiological methods are useful in obtaining objective measures of clinically and experimentally induced pain and have proven to be successful in characterizing ERPs elicited by painful stimuli (Iannetti, Zambreanu, Cruccu, & Tracey, 2005). Previous studies reporting pain ERPs describe an ERP that consists of a negative wave followed by a large positive wave that occurs ca 400 ms after pain onset (Iannetti et al., 2005; Vossen, van Breukelen, van Os, Hermens, & Lousberg, 2011). This positive peak has been labeled differently by several studies, for instance as a P2 (Iannetti et al., 2005) or as a P3 (Vossen, Van Breukelen, Hermens, Van Os, & Lousberg, 2011). In addition, this late positive component has been reported to be increased when the subjective pain experience is more intense and is generated by the cingulate gyrus (Iannetti et al., 2005). Therefore, it has been proposed that this component can be used as an objective measure of experienced pain (Bromm, 1995; Chen, Chapman, & Harkins, 1979).

4.1.4 The present study

The present study investigated pain- and empathy-related neuronal responses in a socially interactive setup. The main aim of this experiment was to investigate the differences in neuronal responses with respect to the participants' role, when observing someone in pain (villain) or receiving a painful stimulus (victim) and their capacity, when actively controlling the painful stimulus (active) or having no control over the painful stimulus (passive). Therefore, we designed a paradigm that included four conditions. During the first condition ('passive villain') the participant passively watched another person pressing a button. During the second condition ('active villain') the participant had to press the button him- or herself. During the third condition ('active victim') the participant received the electrical shocks after pressing the button him- or herself and during the fourth condition ('passive victim') another person was pressing the button while the participant was receiving the electrical shocks. In addition, we asked participants to fill in a self-report questionnaire to measure psychopathic traits in order to investigate the role of psychopathic personality traits on pain- and empathy-related neuronal responses.

We studied four contrasts in this paradigm. The first contrast compared the ERN of the response-locked ERP of the active villain versus the active victim. This enabled us to study the amount of conflict the participant is experiencing when hurting himself or another person. Based on the fact that humans are social animals and are capable of showing empathy towards others, we expected the 'active villain' to show an increased ERN compared to the 'active victim'. Since psychopaths are characterized with low empathy, we expected that this effect correlated negatively with psychopathic traits.

The second contrast considered the potential difference between passive and active observing of another person in pain. The visual P3 component of the visual ERP of the passive villain versus active villain were compared. We expected a higher visual P3 component for the active villain compared to the passive villain condition, since the active role creates a more involved and responsible position for the villain. In line, we expected a negative correlation of psychopathic traits-with the magnitude of the visual P3 effect.

The third contrast compared the P3 component of the visual ERP of the active victim versus the passive victim. This enabled us to study the role of having control over pain. Losing control over a threatening situation increases attention/vigilance which results in an increased visual P3 component. Therefore we expected to find higher visual P3 components for the passive victim compared to the active victim. We did not expect this contrast to be linked to psychopathic traits.

Also the fourth contrast is related to control over pain. We compared the late positive pain component of the pain ERP of the active victim versus the passive victim. It has been demonstrated that pain is perceived as less intense when it is self-controlled (Pellino & Ward, 1998). This effect is reflected in attenuated neural responses in reaction to self-controlled pain (Salomons et al., 2004). Therefore, we expected to find an increased late positive pain component for the passive victim compared to the active victim. We did not expect this contrast to be linked to psychopathic traits.

Social neurocognition is a relatively new emerging field of social cognitive neuroscience. First, the current study provides insight into the influence of social

context, attention and control over pain in self and others. Second, it enables us to better understand the role of psychopathic personality traits on social neurocognition. Third, the paradigm that was designed for this study provides as an alternative, more realistic method to study pain- and empathy-related behaviours.

4.2 Methods

4.2.1 Participants

A total of 60 healthy volunteers (31 females) with an age between 18 and 56 (*M*=31.57, *SD*=8.21) participated during a science fair: The Discovery Festival in Science Centre NEMO, Amsterdam, The Netherlands in September 2015. Before actual participation in the experiment, participants were subjected to a test trial to introduce the nociceptive electrical stimulus. Participants that signed up for the study provided written informed consent and for each participant a short medical checklist was filled out by the researcher. Procedures were approved by the Ethics Committee Social Sciences (registered under amendment ECG2012-1301-010a2) of the Radboud University Nijmegen, The Netherlands. Participants did not receive compensation for participation in this study and participants could leave the experiment at any time. One participant did not complete the whole experiment due to oversensitivity to the stimulation and two participants only completed two out of four conditions of the experiment. In addition, the EEG data of two participants contained excessive artefacts. Data of these five participants were excluded. The data of the remaining 55 participants (29 female; 9 left handed; age *M*=31.80, *SD*=8.00) were further analyzed.

4.2.2 Self-Report Psychopathy Short-Form (SRP-SF)

Before the start of the ERP experiment, participants were asked to fill out the Self-Report Psychopathy Short Form (SRP-SF). The Self-Report Psychopathy (SRP) scale is designed to assess psychopathic traits in an adult non-forensic sample (Hare, 1985). The present study used a Dutch translation of the short version of the SRP (SRP-SF) that included 29 of the 64 original questions. The SRP-SF is highly correlated (r=0.92) with the full version SRP (Paulhus, Neumann, & Hare, 2009) and has been proven to be valid and invariant across gender (Neumann & Pardini, 2014a). The SRP-SF consists of two factors with each two subscales. Factor 1 (F1) covers interpersonal manipulation (e.g. 'Sometimes you need to pretend that you like someone to get what you want') and affective callousness (e.g. 'Most people are weak') and Factor 2 (F2) covers erratic lifestyle (e.g. 'I've often done dangerous things just for the thrill') and overt antisociality (e.g. 'Sometimes I carry a weapon

4

(knife or gun) to protect myself'). Questions needed to be rated on a 5-point Likert scale (1=strongly disagree, 5= strongly agree).

4.2.3 ERP paradigm

Two participants were involved in the task at the same time and EEG of both participants was recorded during the whole experiment. The experiment included four conditions, each consisting of 15 trials. During the first condition ('passive villain') participant N and participant N-1 were seated next to each other while facing the same computer screen. Participant N-1 was instructed to press a large red button which, after 750ms, led to a 200ms-presentation of a visual stimulus (white circle on a black background). The nociceptive electrical stimulus was delivered 750 ms after the onset of the visual stimulus to the left hand of participant N-1 (Figure 4.1).

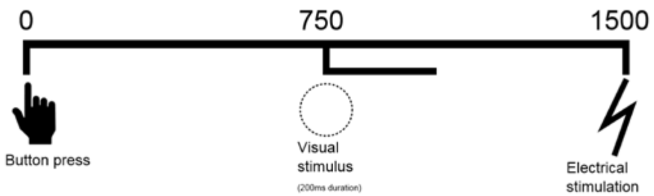


Figure 4.1 Schematic representation of sequence of events. A button press is followed by a visual stimulus on the screen (750 ms) for 200 ms and an electrical shock (1500 ms).

During the second condition of the experiment ('active villain') the roles for pushing the button were switched. Participant N was instructed to press the button while Participant N-1 still received the electrical stimulus. In third condition ('active victim') participant N was moved to the location of participant N-1 who would now leave the experiment. A new participant, Participant N+1, was introduced in the experiment starting with condition 1. Participant N was instructed to press the button which, after stimulus presentation, resulted in the electrical stimulus at his/her own arm while Participant N+1 was observing. In the last condition ('passive victim') participant N and participant N+1 switched roles for pressing the button. Participant N+1 was instructed to press the button while Participant N received the electrical stimulus. For a schematic representation of the design, see Figure 4.2.

Thus, each participant completed all four conditions. We chose not to randomize the different conditions, since, in line with the shared representation model (Decety & Jackson, 2004), previous pain experience or observation of pain

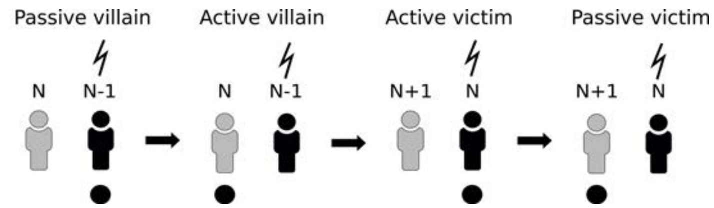


Figure 4.2 Schematic representation of the paradigm. Participant N undergoes all four conditions. In the first two conditions it acts as villain and then switches to victim, which is accompanied by N - 1 leaving the task and N + 1 entering.

in others could influence later pain experience or pain observation in others (Meng et al., 2013; Meng, Butterworth, Malecaze, & Calvas, 2012).

4.2.4 EEG recordings

All measurements were obtained using two mobile EEG labs. EEG and electro-oculography (EOG) signals were recorded with an actiCap-system which uses active Ag/AgCl electrodes (Brain Products GmbH, Munich, Germany). The Fz, Cz, and Pz electrodes were placed according to the international 10-20 system, with an additional electrode on the right mastoid bone, the ground electrode at AFz, and the reference electrode over the left mastoid bone with self-adhesive rings. Post-recording, the electrodes were rereferenced to linked mastoids and filtered between 0.1 and 30 Hz. Electrode impedance was kept below 20 kΩ which is appropriate for active electrodes (Mathewson, Harrison, & Kizuk, 2017). Eye movements were recorded by electrodes placed below the left eye and at the outer canthus of the left eye. The signal was digitized at 1000 Hz.

4.2.5 Stimuli

The response-locked ERNs were captured when a large red button was pressed (diameter: 9.5 cm; height: 5.5 cm), visual ERPs were time-locked to the presentation of a visual warning stimulus (white circle on a black background) for 200ms and pain ERPs were elicited by electric stimuli.

The electrical stimulation was delivered on the volar side of the non-dominant forearm by a concentric ring-electrode (Katsarava et al., 2006) attached to a Digitimer DS7-AH electrical stimulator (Digitimer Ltd). The participant received in total 30 electrical stimuli across both victim conditions, where each stimulus consisted of a rapid train of seven pulses with a 2ms duration and a 2ms inter-pulse-interval. Stimulus intensity was set to correspond to a perceived intensity of 7 on a scale of 0 to 10, where '0' corresponds with "I don't feel anything", and '10' corresponds with “maximum tolerated pain”) beforehand and was kept consistent throughout the experiment. Participants were exposed to short series of test stimuli after which they decided to participate in our experiment. All participants included in the analysis tolerated the painful stimulation.

4.2.6 EEG analyses

The segments belonging to the response, the visual stimulus and the nociceptive stimulus were selected offline. Epochs were defined as ranging from -250ms to 750ms based on stimulus or response markers for each of the three events. Baseline correction was applied using the interval of -250ms to 0ms. To allow blind scoring, component amplitudes were defined as the averaged value within a fixed latency window: The ERN component (20ms-70ms), the visual P3 component (410ms-460ms), the late positive pain $P_{400-500}$ component (400ms-500ms). After visual inspection of the grand average ERPs, the ERN, the visual P3 and pain $P_{400-500}$ could be identified. Amplitudes of these components were determined as the average value within a fixed latency window (Picton et al., 2000). Segments were corrected for EOG artefacts by employing the Gratton & Coles algorithm (Woltering, Bazargani, & Liu, 2013). In contrast to Woltering and colleagues (2013), averaging subtraction was not applied in the current analysis, which left the ERP components of interest unaffected. Trials contaminated with artefacts exceeding 150 µV were excluded. From the total amount of 825 trials that were measured during the experiment, 797 trials were included for further analysis. A 250ms interval was used for baseline correction and response-locked, visual and pain ERPs were subsequently averaged per stimulus type. By averaging, all relevant ERP components were extracted from the ongoing signal according to Table 4.1.

4.2.7 Statistical analyses

The ERN, the visual P3, and the pain $P_{400-500}$ component amplitudes at Fz, Cz, and Pz were further analysed using repeated measures GLMs. The ERN and the pain $P_{400-500}$ were analysed using a 2-by-3 design; capacity (passive/active) or role (villain/victim) and electrode site (Fz, Cz, and Pz) functioned as within-subject variables. The visual P3 was analysed using a 2-by-2-by-3 design, as all four conditions were included, which cover two potential roles (villain/victim) in an active as well as a passive capacity (also see Table 4.1). Greenhouse-Geisser correction was applied when the sphericity assumption was violated. The significance level was set at $\alpha < .05$. Since the hypotheses concerning the contrasts were formulated a priori, no correction of the p-values was required.

Furthermore, we studied the correlations of the total scores and the subscales of the SRP with the difference scores of the contrasts. Difference scores of the contrasts were calculated by subtracting the control condition (passive villain/active victim) from the experimental condition (active villain/passive victim). All statistical analyses were performed in IBM SPSS Statistics Version 22.

Table 4.1 Schematic representation of the conditions, the event-related potentials and the contrasts

	Passive villain	Active villain	Active victim	Passive victim
Motor Response		ERN	ERN	
Interval (ms)		20-70	20-70	
Visual stimulus	P300	P300	P300	P300
Interval (ms)	410-460	410-460	410-460	410-460
Electrical shock			$P_{400-500}$	$P_{400-500}$
Interval (ms)			400-500	400-500

Notes. Horizontal: conditions, Vertical: event-related potentials.

4.3 Results

Grand average response-locked, visual and pain ERPs were constructed (Figure 4.3). An average of 4.1% (*SD*=2.8%) of trials was excluded due to contamination with artifacts.

4.3.1 The ERN component of the response-locked ERPs

The first contrast compared the ERN of the active villain and the active victim (Figure 4.4). A significant main effect for role was found where the ERN was more negative for the villain than the victim ($F(1,54)=6.15$; $p=.016$; $\eta_p^2=.102$) (Figure 4.3: Response ERP). As expected, a main effect for electrode was found ($F(1.53,82.83)=13.19$; $p<.001$; $\eta_p^2=.196$). No interaction effect was observed between role and electrode ($F(1.61,86.83)=1.534$; $p=.223$; $\eta_p^2=.028$). The ERN was maximal at the Fz electrode, therefore the magnitude of the ERN difference at Fz was used to calculate the correlations with psychopathic traits. However, there were no significant correlations with psychopathic traits.

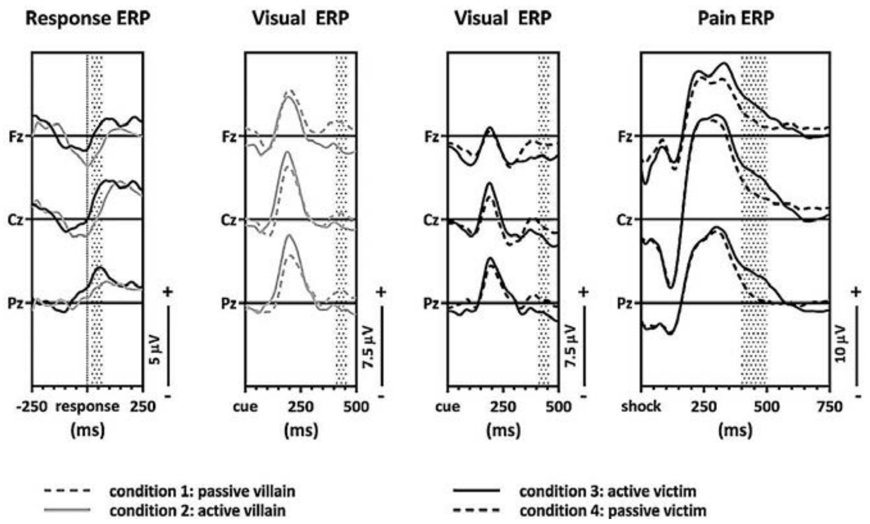


Figure 4.3 Grand average event-related potentials (ERPs). The response-locked ERPs of the active villain vs. active victim, the visual ERPs of the passive villain vs. active villain, the visual ERPs of the active victim vs. passive victim and pain ERPs of the active victim vs. passive victim.

4.3.2 The P3 component of the visual ERPs

Both the second and third contrasts were tested using a single (2-by-2-by-3) overall GLM, which showed an effect of role (villain/victim) (F(1,54)=5.446; p=.023; η_p^2=.092) as well as an effect of capacity (active/passive) (F(1,54)=9.223; p=.004; η_p^2=.146) on the P3. No interaction between role and capacity was present (F(1,54)= .794; p=.377; η_p^2=.014), nor was a three-way interaction apparent (F(2,108)=1.031; p=.360; η_p^2=.019). This means that the effect of capacity was present in both roles, which then relates directly to the two contrasts (contrast two and contrast three). A significant effect of electrode was found (F(1.76,95.22)=3.611; p=.030; η_p^2=.063), with the P3 being maximal at Pz. Therefore, Pz was used for further analysis. There was no interaction effect of electrode with capacity (F(1.77,95.69)=1.345; p=.265; η_p^2=.024) or role (F(2,108)=.804; p=.450; η_p^2=.015).

The second contrast compared the visual P3 of the passive and the active villain (Figure 4.5). A main effect of capacity was found for the P3, as showed by the overall analysis given above. The visual P3 was decreased for the active villain compared to the passive villain (Figure 4.3: Visual ERP). There were no significant correlations with psychopathic traits.

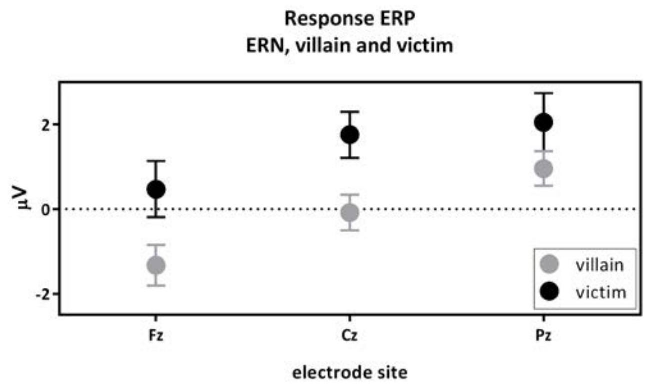


Figure 4.4 The response-locked error-related negativity (ERN) active villain vs. active victim. ERN amplitude (µV) of the midline electrodes (Fz, Cz, Pz) are displayed. The ERN was significantly more negative for the villain than the victim (η_p^2= 0.102; p = 0.016).

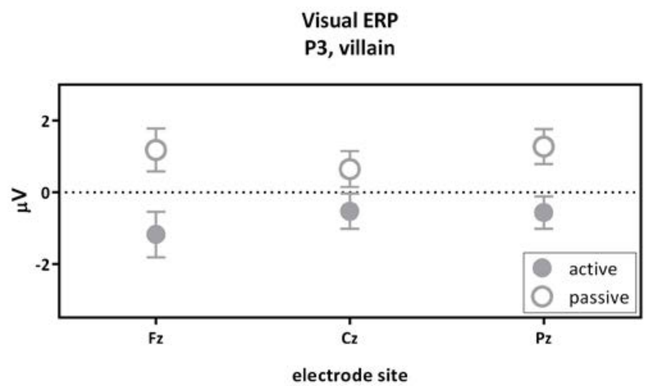


Figure 4.5 The visual P3 passive villain vs. active villain. Visual P3 amplitude (µV) of the midline electrodes (Fz, Cz, Pz) are displayed. The visual P3 was decreased for the active villain compared to the passive villain (η_p^2= 0.116; p = 0.010).

The third contrast compared the visual P3 of the active victim and the passive victim (Figure 4.6). According to the overall GLM, a main effect for capacity was found. The visual P3 was decreased for the active victim compared to the passive victim (Figure 4.3: Visual ERP). There were no significant correlations with psychopathic traits.

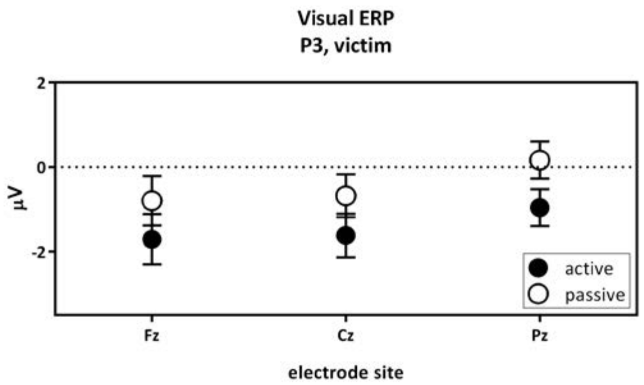


Figure 4.6 The visual P3 passive victim vs. active victim. Visual P3 amplitude (µV) of the midline electrodes (Fz, Cz, Pz) are displayed. The visual P3 was decreased for the active victim compared to the passive victim.

4.3.3 The $P_{400\text{-}500}$ component of the pain ERPs

The fourth contrast compared the pain $P_{400\text{-}500}$ of the active victim and passive victim (Figure 4.7a). There was a significant main effect for capacity (F(1,54)=4.87; p=.033; η_p^2=.082) where the pain $P_{400\text{-}500}$ was decreased for the passive victim compared to the active victim (Figure 4.3: Pain ERP). Moreover, there was a main effect for electrode (F(1.36,73.41)=7.59; p=.004; η_p^2=.123). The pain $P_{400\text{-}500}$ was maximal at the Pz electrode, therefore Pz was used for further analysis. There was no interaction effect for electrode and capacity (F(1.42,76.41)=.473; p=.625; η_p^2=.009) .

The difference score of the pain $P_{400\text{-}500}$ was negatively correlated with the total score on the SRP (r=-.370; p=.005; Figure 4.7b). More specifically F1 scores of the SRP negatively correlated with difference scores of the pain $P_{400\text{-}500}$ (r=-.328; p=.015) and the interpersonal subscale seemed to play an important role (r=-.321; p=.017). F2 scores of the SRP negatively correlated with difference scores of the pain $P_{400\text{-}500}$ (r=-.343; p=.010), where the lifestyle subscale seemed to play an important role (r=-.412; p=.002).

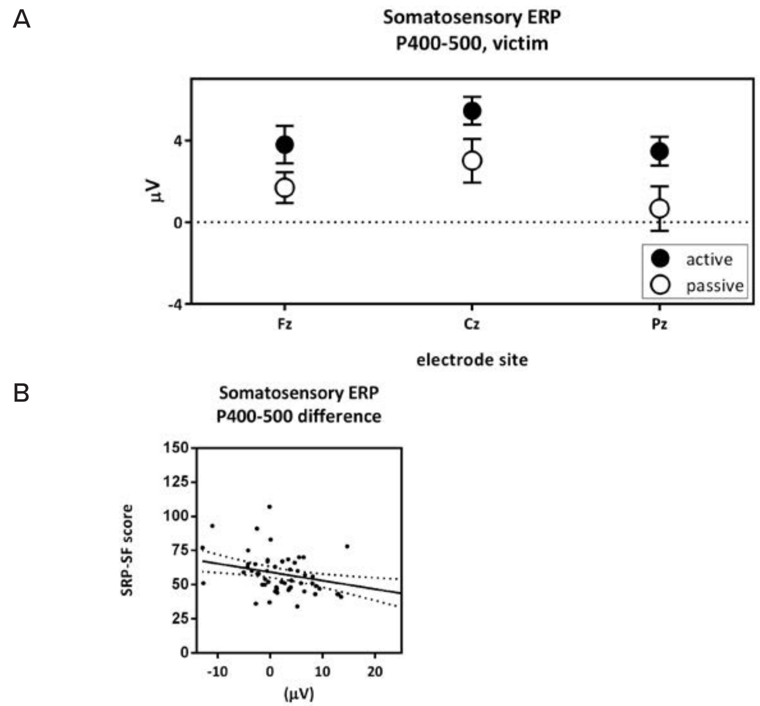


Figure 4.7 Pain $P_{400-500}$ passive victim vs. active victim. $P_{400-500}$ amplitude (µV) of the midline electrodes (Fz, Cz, Pz) are displayed in (**A**). The pain $P_{400-500}$ was decreased for the passive victim compared to the active victim (η_p^2= 0.082; p= 0.033). Scatterplots of the correlations with the total Self-Report Psychopathy (SRP) score (r = 0.370; p < 0.001) are shown in (**B**).

4.4 Discussion

The current study investigated the neural responses of pain- and empathy-related processes in a social, EEG-coupled paradigm. Moreover, we were interested in possible links with psychopathic traits. The first contrast compared the ERNs resulting from the button press of the active villain and the active victim. A clear ERN appeared in response to inflicting pain. Since the ERN is related to conflict, this finding suggests that people experience conflict when hurting themselves or someone else. This outcome is consistent with other findings on empathy (Avenanti, Bueti, Galati, & Aglioti, 2005; Bufalari et al., 2007; Decety, 2010; Singer

& Lamm, 2009). More specifically, our findings suggest that people experience more conflict when hurting someone else than when hurting themselves. To the best of our knowledge, this direct comparison between self-compassion and empathy for others was not made before. Although we would have expected that psychopathic traits correlate negatively with the ERN, we could not confirm this in the current study.

The second contrast compared the visual P3 components of the villain between the passive and active condition. Results indicated a higher visual P3 amplitude for the passive villain compared to the active villain. Previous findings suggested the amplitude of visual P3 to be larger for relevant stimuli than irrelevant stimuli (Steffensen et al., 2008). In the present study the visual stimulus predicting an upcoming shock seems more relevant for the passive villain than for the active villain. For the active villain, the button press already provides information about the upcoming electrical stimulus and the visual stimulus does not add any new information. For the passive villain, the stimulus provides new information. Therefore, we could conclude that our finding is in line with previous literature. This effect did not seem to be influenced by psychopathic traits as no significant correlations were observed.

The third contrast compared the visual P3 components of the active and the passive victim. As expected, results indicated an increase in visual P3 amplitude for the passive victim compared to the active victim. This suggested that the loss of control over the shock led to heightened attention or vigilance in the passive condition. This effect did not seem to be influenced by psychopathic traits as no significant correlations were observed.

The fourth contrast compared the pain $P_{400\text{-}500}$ components of the active and the passive victim. Results showed an increased pain $P_{400\text{-}500}$ amplitude for the active victim compared to the passive victim in response to the shock. This finding is contradicting other studies that suggest that self-controlled pain is perceived as less intense (Pellino & Ward, 1998). A more recent study showed that less predictable pain has a larger impact. However, this is not expressed in pain experience but in physiological impact (heart rate, reaction times) and primary tasks (Arntz & Hopmans, 1998). Moreover, pain literature suggests that attention to pain increases the perceived pain intensity (Villemure & Bushnell, 2009). Actively attributing pain to oneself could heighten attention during the trial, thus also for receiving the shock, and therefore explain the increased pain $P_{400\text{-}500}$ in the active victim in the current study. Besides, the SRP negatively correlated with the pain $P_{400\text{-}500}$ difference score. This suggests that the more psychopathic traits, the less different the pain is experienced in a situation in which the shock is delivered by themselves compared to a situation in which the shock is delivered by another person. Possibly, painful stimuli might be perceived as being less

salient for people scoring high compared to people scoring low on psychopathic traits, and painful stimuli might therefore attract less attention with increased psychopathic traits in both conditions.

In all, recent findings indicate that people experience more conflict when hurting others than when hurting themselves. Furthermore, the results indicate that self-controlled pain was experienced as more painful than uncontrolled pain, which contradicts earlier findings in pain research. Besides, findings showed that people that scored high on psychopathic traits seemed to attend to and experience pain differently. Based on these findings, we suggest that social context, attention and personality traits are important modulators of pain- and empathy-related neuronal responses. Pain experience can be modulated by attention and the way that pain is controlled (self or other). Relevance of being in control or not, the processing of pain predicting stimuli, the salience of such stimuli and attention directed towards these stimuli are all important modulators of empathy-related neuronal responses. In line, psychopathic traits, and indirectly empathic traits, affect pain related neuronal responses.

When interpreting the results, we should take into account that the sequence of conditions was equal for all participants based on ethical considerations. We encountered the order effect of first undergoing the villain conditions followed by the victim conditions. Two distinct forms of perspective taking are described as the ‘imagine-other’ and ‘imagine-self’ perspective. Where ‘imagine-other’ perspective describes the situation in which someone imagines how the other perceives a certain situation and how the other feels as a result, the ‘image-self’ perspective describes the situation in which you imagine how you would perceive a certain situation, were you in the other’s position and how you would feel as a result (Batson, Early, & Salvarani, 1997). The present study measures the imagine-self perspective during the villain conditions, since the villain is aware of the fact that he will be put in the position of the victim afterwards. This could be beneficial because participants experience feelings of distress during the villain conditions (Batson et al., 1997) and this may lead to stronger effects during all conditions. However, the order effect might also be seen as a limitation. Since the active victim condition is always first, this could result in a habituation effect for the passive victim.

Another limitation is that the inclusion criteria were lenient. For instance, age was not restricted and from previous literature we learned that older participants show longer P3 latencies (Kuba et al., 2012; Mullis, Holcomb, Diner, & Dykman, 1985). Moreover, the experiment was done overnight at a science festival. These limitations were mostly controlled by the within-subject design and even though this experiment was performed in a semi-controlled environment, we found robust effects that were overall in line with previous literature.

All in all, we suggest that both pain and empathy-related neuronal responses are modulated by social context, attention and personality traits. Moreover, the within-subject experimental design described in this study thus provides an excellent approach to further unravel the influence of personality traits on social cognition.

5

Dissecting how psychopathic traits are linked to disturbed learning from reward and social advice

This chapter is based on:
Driessen, J.M.A., Diaconescu, A., Buitelaar, J. K., Kessels, R. P. C., Glennon, J. C., & Brazil, I. A. (in preparation). Dissecting how psychopathic traits are linked to disturbed learning from reward and social advice.

Abstract

Individuals that are characterized with a high level of psychopathic traits tend to consistently violate social norms and seek personal gains. Previous studies indicated that this behavior could be caused by impairments in associative learning. In order for associative learning to be successful, we need to constantly monitor the relationship between events and their outcomes and adapt our behavior in response to changes in these event-outcome contingencies. Recent advances in computational modelling approaches offer insight into the unobservable computational processes that are thought to be required for associative learning. In the current study, we used such a model to investigate the associations between psychopathic traits and the computational processes underlying adaptive behavior during associative learning. Moreover, we were interested in the role of oscillatory theta activity, given its potential involvement in adaptive control processes. Participants performed a reinforcement learning task in which the trade-off between using social and reward (non-social) information affected the task performance and the associated monetary reward for the participant. The findings indicated that increasing levels of psychopathic traits co-occurred with greater impairments in learning from social information, and suggested that interpersonal traits were linked to a reduced ability to adapt to changes in the trustworthiness of social advice. These impairments did not lead to a preference for one of the information sources, and the decreased task performance did not affect the risk that was taken in order to obtain a high reward. Furthermore, it was found that decreased theta power was linked to higher levels of psychopathic traits, which aligns with indications that theta is involved in tracking the volatility of social information. These findings contribute to our understanding of how psychopathic traits are associated to adaptive mechanisms underlying associative learning.

5.1 Introduction

Psychopathy is a personality construct that is typified by social-emotional deficits, an erratic lifestyle, and antisocial behavior (Hare, 1985, 1991). It has been demonstrated that the construct of psychopathy is dimensional in nature and consists of a constellation of traits that vary along a continuum (Gao & Raine, 2010; Hare & Neumann, 2005; Hare & Neumann, 2008; Levenson et al., 1995). This line of work has shown that psychopathic traits can be measured in both offenders and among the general community (Brandt, Kennedy, Patrick, & Curtin, 1997; DeMatteo, Heilbrun, & Marczyk, 2006; Malterer, Lilienfeld, Neumann, & Newman, 2010; Neumann & Hare, 2008). The historically dominant view postulates that psychopathy consists of four facets capturing disturbances in the interpersonal (e.g. showing superficial charm, manipulative behavior and deceitfulness) and affective domains (e.g. being callous, lack of empathy and guilt), combined with the tendency to lead and erratic lifestyle (e.g. impulsivity, leading a parasitic lifestyle, irresponsibility) and to engage in antisocial behaviors (e.g. aggressiveness, juvenile delinquency and criminal versatility). The combination of interpersonal and affective facets is believed to be unique to psychopathy, while the lifestyle and antisocial components represent more general traits that can be found across antisocial populations (Blackburn, 2007; Hare & Neumann, 2009). Support for this model has been provided by several studies across a wide variety of samples (Hill et al., 2004; Kosson et al., 2002; Neumann et al., 2007; Olver et al., 2013).

In clinical settings, individuals with elevated levels of psychopathy are known for their reduced ability to monitor and adequately adapt their behavior in response to treatment (Ly et al., 2016). Research has shown that these maladaptive tendencies are linked to impairments in associative learning (Blair et al., 2004; Mitchell, Richell, Leonard, & Blair, 2006; Newman & Kosson, 1986; Patterson & Newman, 1993). For instance, psychopathic offenders show deficits in learning about the association between a stimulus and a reinforcer that occurs in temporal proximity to this stimulus (i.e. contingency learning) (Brazil et al., 2013; Brazil, Maes et al., 2013; Gregory et al., 2015; Von Borries et al., 2010). Associative learning also plays a key role in the acquisition of desirable social behaviors (Behrens, Hunt, & Rushworth, 2009) and seems to be one of the mechanisms through which we learn about social expectations and moral reasoning (Blair, 2013; Blair & Cipolotti, 2000). The tendency to consistently violate social norms and to behave antisocially clearly indicates that there are disturbances in social learning in psychopathy (Brazil et al., 2011). Indeed, individuals with psychopathic tendencies have been found to show impaired learning from social information (Blair, 2007; Brazil et al., 2013; Brazil et al., 2011) and fail to take advantage of standard socialization techniques (Wootton, Frick, Shelton, & Silverthorn, 1997). As a consequence, they

5

are more likely to learn to use antisocial (moral) strategies to achieve their goals (Driessen, van Baar, Sanfey, Glennon, & Brazil, in press; chapter 6), which has a detrimental impact on their social environment and their acceptance by their peer group.

Recent studies have provided insight into the computational mechanisms that are thought to underlie associative learning (Behrens, Hunt, Woolrich, & Rushworth, 2008; Brazil et al., 2013; Cohen, 2008; Oba et al., 2019; Sutton & Barto, 2011). Learning the value of information requires the accurate generation of estimates (or 'representations') of the characteristics of each stimulus-outcome contingency (Brazil et al., 2017). As we live in a complex and dynamic environment, these contingencies may change over time (Dayan, Kakade, & Montague, 2000). Consequently, we need to constantly monitor the relationship between events and their outcomes and update our representations of the contingencies after each observation in order to keep the representations accurate. However, we are not capable of forming perfect representations because i) there will always be inaccuracy in how we process incoming sensory information (Findling, Skvortsova, Dromnelle, Palminteri, & Wyart, 2019), and ii) the characteristics of the contingencies could change over time (i.e. volatility) (Behrens, Woolrich, Walton, & Rushworth, 2007). The level of volatility determines how much a person learns; we learn the least when contingencies remain stable and predictable in the environment (i.e. low learning rate and low uncertainty), while rapid changes in the environment force us to learn quickly how we should adapt our estimates of the contingencies (i.e. high learning rate and high uncertainty) (Behrens et al., 2007; Courville, Daw, & Touretzky, 2006; Dayan et al., 2000). Furthermore, the rate at which contingency changes are perceived to occur affects our representation about the overall likelihood that the contingencies will change, indicating that these two quantities are hierarchically coupled (Brazil et al., 2017; Mathys, Daunizeau, Friston, & Stephan, 2011; Mathys et al., 2014). Thus, appropriate adaptation of behavior depends in part on our ability to learn about the rate and overall likelihood with which contingencies can change in our environment, and also on how well these two processes interact.

Importantly, behavioral adaptation can be triggered by changes in social and non-social factors. Previous studies have demonstrated that learning from social information and (non-social) reward are subserved by similar mechanisms, even though these are implemented in distinct neural regions (Behrens et al., 2008; King-Casas et al., 2005). Neuroimaging findings have pointed out that the learning rate predicted BOLD activation in the anterior cingulate sulcus (ACCs) during social learning, whereas the learning rate predicted BOLD activation in the anterior cingulate gyrus (ACCg) during reward-based learning (Behrens et al., 2008). These findings are in line with earlier studies indicating a functional dissociation between ACCs and ACCg for processing of social and non-social information, respectively, in both man (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005) and macaques (Behrens et al., 2007; Matsumoto, Matsumoto, Abe, & Tanaka, 2007; Rudebeck, Buckley, Walton, & Rushworth, 2006; Van Hoesen, Morecraft, & Vogt, 1993).

After the information related to volatility has been coded by the substructures of the ACC, the information is then combined in the medial prefrontal cortex (mPFC) to guide further adaptation of behavior (Behrens et al., 2008). Electrophysiological findings have suggested that midfrontal activation in the theta frequency band, recorded from electrodes over the mPFC, reflects a common mechanism for implementing adaptive control in situations involving uncertainty about actions and outcomes (Cavanagh, Frank, Klein, & Allen, 2010; Cavanagh, Zambrano-Vazquez, & Allen, 2012). Theta activity was found to be associated with representations concerning the occurrence of reward and also with learning rate (Cavanagh, Figueroa, Cohen, & Frank, 2012; Cavanagh et al., 2010; Cavanagh, Zambrano-Vazquez et al., 2012; Cohen, Elger, & Ranganath, 2007; Ferdinand, Mecklinger, Kray, & Gehring, 2012; Marco-Pallares et al., 2008; Mas-Herrero & Marco-Pallarés, 2014; Talmi, Atkinson, & El-Deredy, 2013). For instance, Mas-Herrero and Marco-Pallarés (2014) demonstrated that theta-band activation is linked to inaccuracies in expectations of reward (i.e. reward prediction errors) in both the acquisition of contingencies and after reversal of these contingencies. Furthermore, they found that fluctuations in theta band activity covaried with the learning rate across participants during both types of learning. Theta activation was larger during reversal learning as compared to acquisition learning, which is in line with the suggestion that the mPFC, including the anterior cingulate cortex (ACC), tracks the environmental volatility (Behrens et al., 2007). The authors proposed that increases in frontal theta activity could be analogous to the increase of mPFC activity associated with the learning rate that were demonstrated in previous fMRI studies (Behrens et al., 2007; Jocham, Neumann, Klein, Danielmeier, & Ullsperger, 2009; Krugel, Biele, Mohr, Li, & Heekeren, 2009; Walton, Devlin, & Rushworth, 2004; Yoshida & Ishii, 2006). Thus, midfrontal theta can be seen as an electrophysiological marker reflecting adaptive control processes during associative learning.

Although research on EEG dynamics in the time-frequency domain in relation to psychopathy is scarce (Clark, Bontemps, Batky, Watts, & Salekin, 2019), a few studies have indicated a role for the theta frequency band in information processing in psychopathic offenders. Two studies found reduced theta in response to affective stimuli (Eisenbarth et al., 2013; Tillem et al., 2016) and this was suggested to underlie impairments in the processing and integration of sensory information (Tillem et al., 2016). Other work has examined theta in relation to external correlates of psychopathy, such as externalizing behavior (Bernat et al., 2011) and antisocial behavior (Mednick, Vka, Gabriella, & Itil, 1981). Thus, there is no clear consensus on the role of theta oscillations in psychopathy. Moreover, there are no studies that looked into adaptive control processes and theta in relation to psychopathic traits.

The link between psychopathy and disturbances in associative learning was initiated by studies using basic behavioral outcomes such as accuracy rates and reaction times (e.g. Blair, 2004; Brazil, Maes, et al., 2013; Gregory et al., 2015; Mitchell et al., 2006; Newman & Kosson, 1986; Patterson & Newman, 1993; Von Borries et al., 2010). Although these studies provided valuable insights, the approach employed did not allow a direct quantification of the different latent computations occurring within the mechanism of associative learning. Recent advances in computational modelling approaches offer insight into the unobservable computational processes that are thought to be required for associative learning, such as estimations of changes in reward probability or reward size (e.g. Behrens et al., 2008; Behrens et al., 2007; Mathys et al., 2014). By examining associative learning in a group of individuals with varying levels of psychopathic traits with the use of such a model, we could obtain deeper insight into the latent processes within the associative learning mechanisms that are disturbed these individuals. To date, only a few studies have employed methods that allow for a more detailed view into the learning mechanism (Blair, 2004; Brazil et al., 2017; Brazil, Maes, et al., 2013; Oba et al., 2019). The increased precision offered by such an approach can be further enhanced by additionally collecting electrophysiological measurements of the mPFC that can be used as direct measures of adaptive control during associative learning. Therefore, the present study combined computational modelling with electrophysiological measurements in order to investigate how levels of psychopathic traits covary with i) the use of social and non-social information during associative learning, ii) the updating of representations concerning the volatility of reward and social advice, and iii) oscillatory theta band activation during associative learning. Based on findings pointing towards impaired associative social learning in psychopathy (Blair, 2013; Brazil et al., 2013; Brazil et al., 2011), we expected interpersonal-affective traits to be negatively related to the use of social advice during learning. In contrast, we predicted that lifestyle-antisocial traits should be positively linked to the use of reward during learning, based on previous studies linking high levels of impulsive-antisocial traits to hypersensitivity to reward (Buckholtz et al., 2010). At the computational level, we hypothesized that increasing levels of interpersonal-affective and antisocial traits should be linked to disturbances in the representation of volatility, based on findings suggesting that these psychopathic features are linked to an excessively large variability in the representation of volatility during threat conditioning (Brazil et al., 2017). Finally, we expected reduced theta activity in

relation to psychopathic traits, based on the proposed role of theta in adaptive control during associative learning and previous studies that indicated a negative link between theta band activation and psychopathy.

5.2 Methods

5.2.1 Participants

Based on a priori samples size calculations (power=.85), eighty-six participants were included in this study (36 males). Participants were between 18 and 35 years of age (M = 24.1, SD = 2.8) and were native Dutch speaking. Exclusion criteria for the current study were self-reported epilepsy, head surgery, claustrophobia, history of other neurological conditions or psychiatric disease, use of psychoactive medication or substances, and pregnancy.

Participants were selected from a large database (N=1519, 309 males) based on their level of psychopathic traits, which was assessed with the short-form version of the self-report psychopathy checklist (SRP-SF; Dutch version: Gordts et al., 2017). The SRP-SF is a self-report questionnaire including 29 items that need to be rated on a 5-point Likert scale. Subscale scores (interpersonal, affective, lifestyle, and antisocial) can be derived as well as a total score. A substantial amount of literature supports the validity and reliability of the SRP-SF in community samples (Gordts et al., 2017; Hare & Neumann, 2008; Lilienfeld et al., 2006; Neumann & Pardini, 2014). Internal consistency in the sample was high (all Cronbach's alpha > .75). Participants in the large database were recruited via local advertisement and an online research participation system (SONA systems) of the Radboud University in Nijmegen. The total scores in the large sample (N=1519, 309 males) were divided in quartiles to make sure that 25% of the participants ended up in the top and bottom quartiles, while 50% of the participants ended up in the two middle quartiles. We selected 86 participants (36 males) based on their SRP total score for the current study. The top and bottom quartiles were oversampled in order to enhance the presence of extreme scores on both sides of the distribution (8, 60). Consequently, 20 participants (23.3%) belonged to the lowest quartile, 40 participants (46.5%) belonged to the two middle segments, and 26 participants (30.2%) belonged to the upper quartile.

The participants provided written informed consent prior to the onset of the study, and received monetary compensation at the end of the study. Procedures were approved by the Ethics Committee Social Sciences of the Radboud University Nijmegen, Netherlands. Six participants showed major distortions in at least one of the midfrontal electrodes, therefore we excluded these participants from the electrophysiological analyses.

5.2.2 Wager task

The Wager task is a modified version of the deception-free binary lottery game created by (Diaconescu et al., 2014) (Figure 5.1). The task included 160 trials with a short break after 70 trials. In each trial, participants were asked to predict the outcome of a card draw – blue or green- while they had access to two sources of information, information from the "advisor" (social information) and individually experienced recent outcome history (reward information). In each trial, the draw was preceded by a short video clip that showed the advisor holding up one of the two cards, therewith recommending to the participant which card to choose. The video clips were previously recorded face-to-face sessions (Diaconescu et al., 2017). The advisors based their suggestion on true but probabilistic information with a constant probability of 80%, however, this was not made aware to the participant. Furthermore, the advisors received monetary incentives to change their strategy and provide either helpful or misleading advice at different phases of the task. Although the original study included four advisors (two males, two females), the current study used the video clips of one male advisors for all participants in

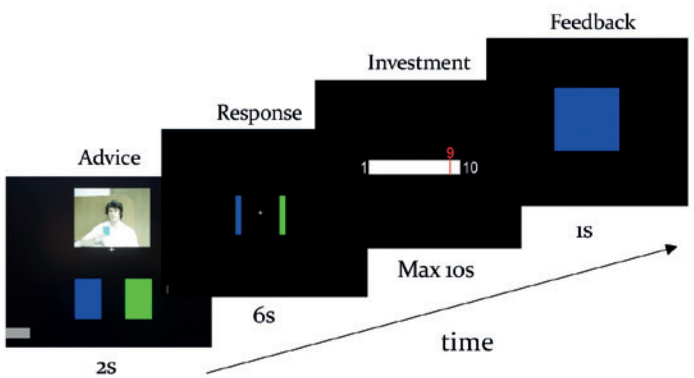


Figure 5.1 Schematic overview of the Wager task. The participant predicts the outcome of a binary lottery in which the color of a card (blue or green) corresponds to a positive or negative reward (points). The participant can use two sources of information; 1) outcomes on previous trials (i.e. reward information), 2) advice of a confederate who is more informed about the outcomes of the task (i.e. social information). After the participant has chosen a card he/she can invest a certain amount of points (0-10) that determines the reward in case the prediction was correct (add points) or incorrect (deduct points). Each trial consists of four phases: 1) Advice: advisor recommends a card, 2) Response: participant choses a card, 3) Investment: participants decide how many points to invest, and 4) Feedback: outcome was presented.

order to reduce variance in the task. (Diaconescu et al., 2017) showed that there were no performance differences or differences in reliance on the advice between different advisors in a similar task.

The color-reward probabilities and the advisor intentions varied independently across trials. This resulted in four different task phases, 1) stable card and volatile reward, 2) stable card and stable reward, 3) volatile card and stable reward, and 4) volatile card and volatile reward (Figure 5.2). This enabled us to investigate the effect of volatility for the card and advice independently.

In addition to the opportunity to infer the advice accuracy as a function of the advisor's intentions, participants could infer the card probabilities based on the reward outcome history. After the selection of one of the cards (blue or green), participants were asked to wager a number of points between one and ten to indicate how confident they were about their predictions. In each trial, the start position of the marker of the wager was random to ensure that participants did not

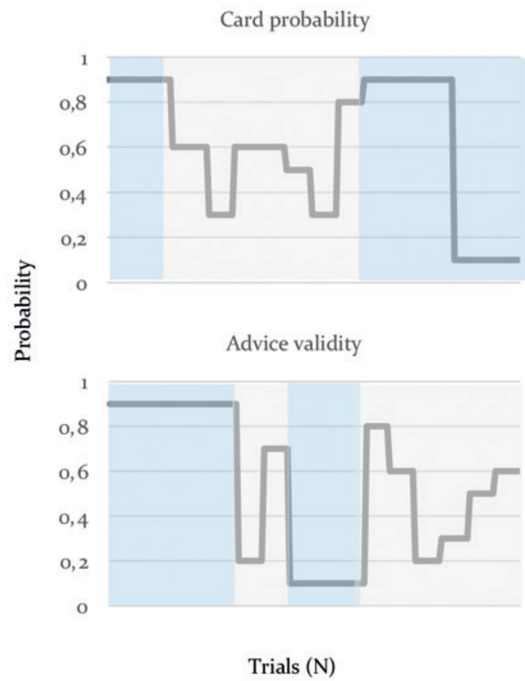


Figure 5.2 Probability scheme for card and advice trials. Stable phases are highlighted in blue (Card: trials 1-25 and 100-160; Advice: trials 1-49 and 70-99), volatile phases are in grey (Card: trials 26-99; Advice: trials 50-69 and 100-159).

wager points based on the previous trial. The score on each trial was dependent on the participant's wagered amount of points and, consequently, the participant's payment at the end of the experiment. After the wager screen, the winning card (i.e. outcome) was presented. For each trial, the cumulative score was updated by the won or loss wager of the previous trial and was presented as a bar at the bottom of the screen. A white bar indicated a positive score and a red bar indicated a negative score. To increase the participants' motivation, they were told that they could reach a silver (€2,50) and a golden (€5,00) bonus target. These targets were invisible during the task in order to avoid loss of motivation. At the end of the task participants were told the amount of bonus that they earned.

5.2.3 Procedure

The present study is part of a larger project in which participants performed three different computer tasks. The order of the tasks was randomized between participants and tasks other than the Wager task are not considered in this chapter. Participants received instructions before the start of the task, and the experimenter checked whether each participant understood the instructions correctly. At the end of the experiment, participants received their monetary compensation and the bonus that was determined by the total amount of points at the end of the task.

5.2.4 Electrophysiological recordings

Electrophysiological data were collected using 32 active scalp electrodes (ActiCap, Brain Products, Munich, Germany) arranged according to a variation of the international 10–20 system, with an additional electrode on the right mastoid. Electrophysiological data was acquired at 500 Hz without filtering with the QuickAmp amplifier (Brain Products), with the reference electrode over the left mastoid bone with self-adhesive rings. In addition, six passive electrodes were used to collect horizontal and vertical eye movements and an electrocardiogram. Vertical eye movements were recorded by placing electrodes above and below the left eye and horizontal eye movements were registered at the outer canthi of the eyes.

5.2.5 Analyses

5.2.5.1 Computational modelling of behavioral responses

In the present study, we examined how subjects updated their beliefs about others' intentions and chose to follow or disregard their advice. For this purpose, we applied a computational model which is referred to as the Hierarchical Gaussian Filter (HGF), a generic Bayesian model of learning under perceptual uncertainty and environmental uncertainty (Mathys, 2012). In order to verify whether participants estimate the volatility of information sources, we also included a reduced version of the HGF (2-level HGF) containing only two levels of learning as control.

The HGF model is widely used for computational analyses of behaviour (e.g. De Berker et al., 2016; Hauser et al., 2014; Iglesias et al., 2013; Marshall et al., 2016; Vossel et al., 2014). In the current analyses, we assumed that the rewarding card colour (reward-based learning) and the advice accuracy (social learning) varied as a function of hierarchically coupled hidden states χ1(k), χ2(k),... χn(k). They evolved in time by performing Gaussian random walks. At any given level, the step size was controlled by the state of the next-higher level (Figure 5.3). The lower level hierarchy states χ1a(k) and χ1c(k) were binary and represented the advice accuracy (1 for accurate, 0 for inaccurate) and the rewarding card colour (1 for blue, 0 for green). The second level hierarchy states χ2a(k) and χ2c(k) were continuous variables and represented the advisor's reliability and the tendency for a given card colour to be rewarding. The third level hierarchy states χ3a(k) and χ3c(k) were also continuous and denoted the rate of change of the advisor's intentions and the card-outcome contingencies. Four learning parameters, κa, κc, ϑa, and ϑc, determined the rate with which the hidden states evolved in time. The parameters κa and κc determined the coupling between the second and third hierarchy levels, whereas ϑa and ϑc determined the volatility over time. The model we describe here is called a generative model and illustrates the process of producing the outcomes that are observed by the participant. In agreement with Bayes' rule, we assume that participants who make inferences on advice and card outcomes, form posterior beliefs over the hidden states (i.e., advice-outcome congruency and the rewarding card colour) based on the inputs they observe. The application of the Bayes' rule to our generative model (i.e. model inversion) results in a perceptual model, which describes participants' beliefs about hidden states.

The response model maps the predicted outcomes to behavioural responses (blue or green card) and predicts two components of the outcome on each trial; 1) the participant's decision to accept (y=1) or reject (y=0) the advice, and 2) the number of points wagered. We assume that a prediction of the outcome on a given trial is defined as a function of arbitration (i.e. the perceived reliability of each information source) and the predictions afforded by each source. Response model parameter ζ represents the participant's bias towards social information, and has a prior mean of 1, indicating an equal weighing of the two information sources. A value of ζ>1 indicates a bias towards advice, while a value of ζ<1 indicates a bias towards the estimated card probability. The formal equations underlying the model and the model inversion are described in Diaconescu et al. (2019).

5.2.5.2 Wavelet analysis

Electrophysiological data was analyzed using MATLAB (v2015b, MathWorks Inc., Natrick MA), in combination with the Fieldtrip toolbox (Oostenveld, Fries, Maris, &

Schoffelen, 2011). Data were filtered offline using a 0.01-40Hz filter and re-referenced to the average of the linked mastoids. Manual artifact rejection was conducted in order to detect muscle and electrode artifacts in the data. Subsequently, EOG and ECG artifacts were removed using the default Independent Component Analysis in the Fieldtrip toolbox. With the use of this ICA, components that contained ocular artifacts were identified by inspecting the time course and spatial topography of all components. Activity associated with the feedback stimulus was averaged separately in epochs starting 500ms prior to the stimulus onset and ending 1500ms after stimulus onset. Next, time-frequency representations (TFRs) were

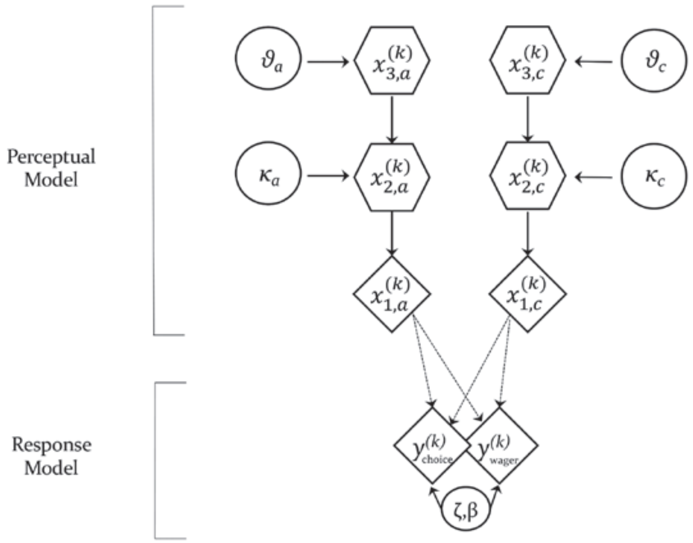


Figure 5.3 Graphical model of the HGF and the response model. The diamonds represent quantities that change in time (i.e., that carry a time (or trial index k)) but that do not depend on their previous state. The hexagons represent states that change in time but additionally depend on their previous state. The circles denote fixed parameters. The perceptual model has three layers: (1) χ1 represents the accuracy of the current advice, (2) χ2 the adviser's current tendency to give accurate advice and (3) χ3 the current volatility of the adviser's intentions. Parameter κ determines how strongly χ2 and χ3 are coupled, and ϑ indicates the meta-volatility in χ3. The response model has 2 layers: (1) the computation of the probability of the outcome given both the non-social cue and the advice; (2) the chosen action. Parameter ζ determines the weight of the advice compared to the non-social cue. Y represents the subject's binary response (y~1: accept the advice, y~0: reject the advice). Adapted from Diaconescu et al. (2017; 2019).

calculated using Morlet wavelets with a width of 3 (width of the wavelet in number of cycles, (Tallon-Baudry & Bertrand, 1999) and a frequency range of 1-40Hz within -0.5 – 1.5 seconds in 40 steps for the different conditions (overall, correct, incorrect, stable card, volatile card, stable advice, and volatile advice) separately. Further analyses focused on the average of the evoked theta power (4-8Hz) over the midline electrodes Fz, FCz, and Cz during a 100-400ms time window after the outcome (blue or green card) was presented on the screen.

5.2.5.3 Statistical analyses

To test for significant effects of phase (stable, volatile) and cue (card, advice) on performance accuracy, advice taking, and wager magnitude, general linear models (GLMs) for repeated measures with phase and cue as within-subject measures were used (two-tailed). To examine whether the predicted wager accurately predicted the actual points wagered by participants, a correlation analysis was performed between the predicted and actual wager.

Model comparison between the non-hierarchical and the hierarchical HGF models was performed with the Bayesian Omnibus Risk (BOR) test. The model with the best fit was used for further analyses with the computational parameters. In order to test the model's internal validity, a linear regression analysis was performed to examine whether the model predictions were consistent with participant's ratings of adviser's trustworthiness during the experiment. Furthermore, Spearman correlations were conducted to investigate whether the wager magnitudes predicted by the model significantly correlated with the participant's actual wagers. Cue effects on the computational learning model parameters κ and ϑ were tested using a repeated measures GLM with cue as within-subject measure. Furthermore, a one-sample t-test was performed to investigate ζ was different from 1. Finally, the link between performance accuracy, advice taking, wager magnitude, and the computational learning parameters and the SRP total and subscale scores were tested using Spearman correlations.

To test for significant effects of phase and cue on mean theta power over the midline electrodes (for short, 'theta power'), a GLM for repeated measures with phase and cue as within-subject variables were used (two-tailed). A paired t-test was used to study a potential difference between theta power over correct and incorrect trials. Moreover, Spearman correlations were conducted to examine the link between theta power and the overall accuracy, the accuracy over volatile and stable trials, and the computational learning model parameters. Lastly, Spearman correlations were tested to study the link between psychopathic personality traits and theta power over the midline and over the midline electrodes separately (Fz, FCz, Cz).

5.3 Results[1]

5.3.1 Behavioral results

5.3.1.1 Performance accuracy, advice taking and wager magnitude

Overall performance accuracy was 60.1% ± 4.1% (mean ± standard deviation). There was a main effect of phase (F(1,85)=358.314, p<.001, η_p^2=.808), and a cue-by-phase interaction (F(1,85)=38.017, p<.001, η_p^2=.309). Post hoc analyses revealed that participants performed significantly better in stable as compared to volatile advice phases of the task (t(85)=11.197, p<.001), while there was no phase difference for card learning (t(85)=.163, p=.871). This indicated that the degree to which participants relied on the advice information source was a function of precision, which in turn, depended on the volatility structure of the task (Figure 5.4a). Participants took advice in 57.6% ± 9.0% of the trials. There was also a main effect of phase (F(1,85)=41.884, p<.001, η_p^2=.330), and a cue-by-phase interaction (F(1,85)=17.741, p<.001, η_p^2=.173) on advice taking. Post hoc analyses showed that participants took advice into account more often in stable as compared to volatile advice phases of the task (t(85)=6.062, p<.001). There was no significant effect of card phase on advice taking (t(85)=-.880, p=.382)(Figure 5.4b). Finally, the overall average points wagered was 5.00 ± 2.91. A main effect of phase on the amount of points wagered was observed (F(1,85)=98.409, p<.001, η_p^2=.537). This indicated that, regardless of the cue type, participants wagered significantly more points during stable phases of the task as compared to volatile phases (Figure 5.4c).

Performance accuracy during volatile phases of the task appeared to be associated with psychopathic traits. Accuracy during volatile trials was negatively correlated with lifestyle traits (r=-.221, p=.041), and antisocial traits (r=-.232, p=.032). This indicated that individuals with high levels of lifestyle-antisocial traits were more affected by volatility. More specifically, the effect of psychopathic traits on accuracy on volatile trials was driven by the volatile advice trials (total score: r=-.235, p=.029), interpersonal: r=-.228, p=.035), affective: r=-.253, p=.019; antisocial: r=-.231, p=.032), while the effect was not present in the volatile card trials (Table 5.1; Bayesian correlations for SRP and performance on volatile advice trials can be found in Table S5.1). Psychopathic traits were not associated with the degree of advice taking or the amount of points that they wagered on the different trial conditions.

1 In addition to the results that followed from frequentist statistics that were reported in the result section, we tested the correlations with Bayesian statistics. These outcomes of the two approaches are consistent, and the Bayesian results can be found in the supplement.

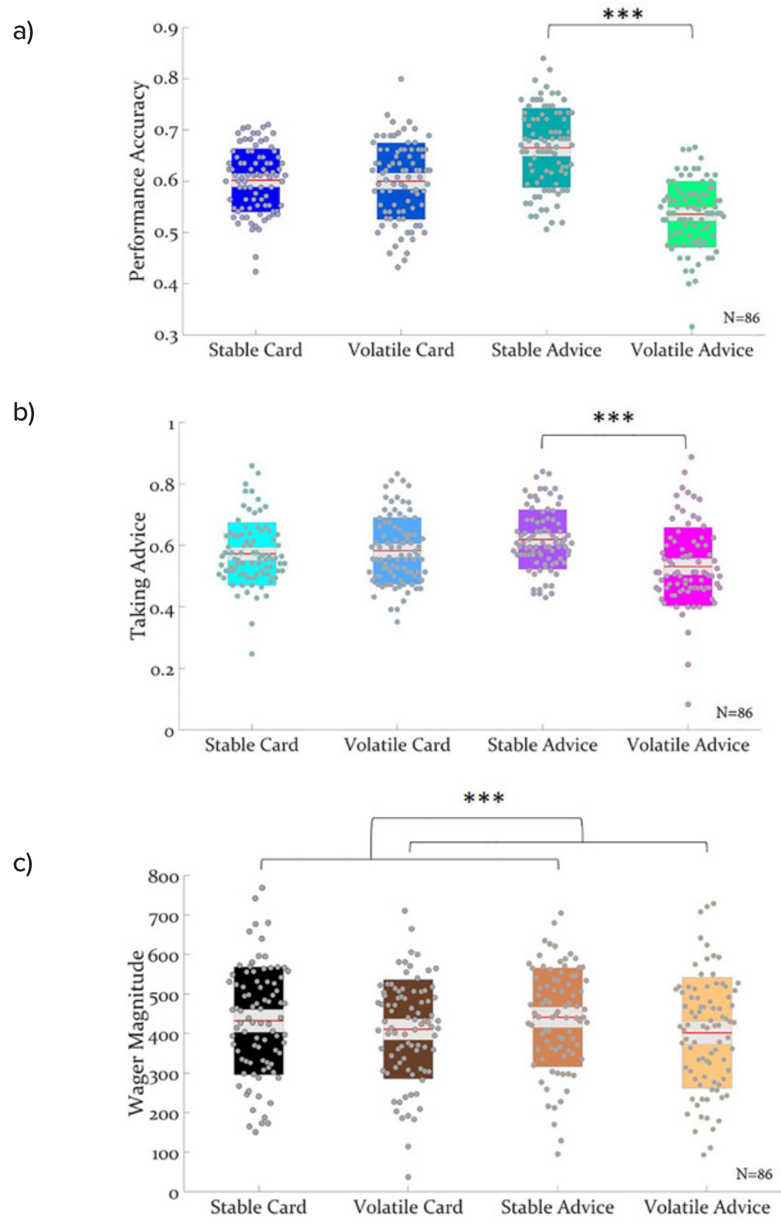


Figure 5.4 The mean of the behavioural variables influenced by the volatility: **a)** performance accuracy, **b)** advice-taking behaviour, and **c)** the amounts of points wagered. The red line refers to the mean, and the grey background reflects the 95% confidence intervals for the mean.

5

Table 5.1 Spearman correlations between SRP scores and behavioral parameters

	SRP Total		SRP Interpersonal		SRP Affective		SRP Lifestyle		SRP Antisocial	
	rho	p	rho	p	rho	p	rho	p	rho	p
All	-.18	.100	-.15	.163	-.13	.222	-.21	.054	-.14	.187
Stable	-.15	.161	-.12	.226	-.11	.316	-.19	.083	-.08	.147
Volatile	-.21	.059	-.18	.085	-.15	.156	-.22*	.041	-.23*	.032
Stable card	-.17	.110	-.11	.217	-.19	.081	-.15	.172	-.11	.322
Volatile card	-.12	.255	-.12	.270	-.05	.682	-.17	.111	-.14	.194
Stable advice	-.09	.390	-.08	.477	-.04	.751	-.15	.164	-.05	.670
Volatile advice	-.24*	.029	-.23*	.035	-.25*	.019	-.16	.154	-.23*	.032

Notes. Asterisks indicate significant correlations (*$p<.05$).

5.3.1.2 Model-based results

5.3.1.2.1 Model selection

Computational modelling of behavior was used to explain participant's responses on every trial. We compared two models that assumed different mechanisms for arbitrating between individual and social learning. Bayesian model selection revealed one winning model. Over all participants we found that the model that assumed that participants tracked volatility of both information sources outperformed the model without volatility tracking (Bayes Omnibus Risk = .708) (Figure 5.5). Based on these findings, we decided to use the outcomes of the 3-level hierarchical model for further analyses.

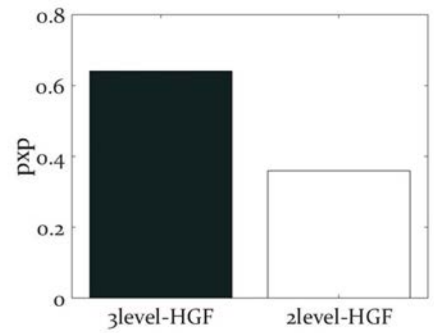
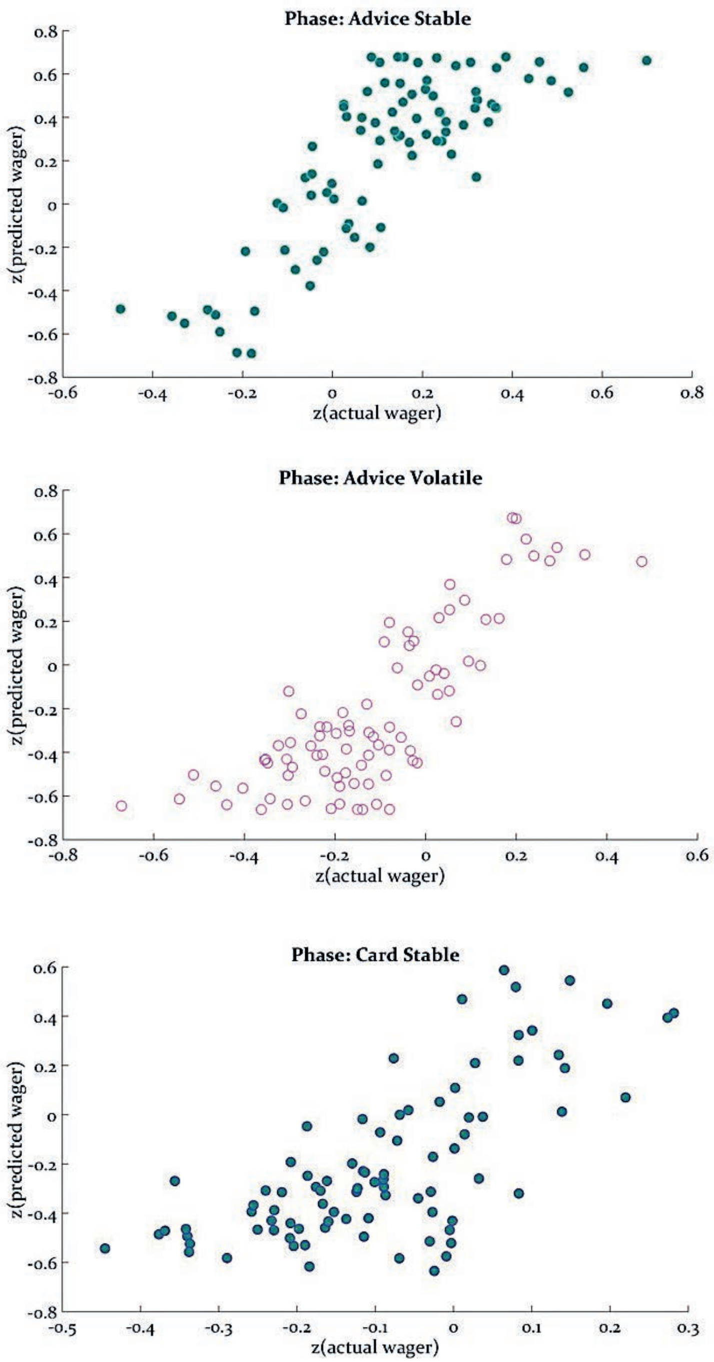


Figure 5.5 Model selection results. Comparison of the 3-level HGF model and the 2-level HGF model (no volatility tracking). The posterior probability for each model given the participants' data obtained using random effects BMS suggest that the 3-level HGF model explained participants' responses better as compared to the 2-level HGF model.

5.3.1.2.2 Model internal validity

As an internal validity check, we examined whether the model predictions were consistent with participant's ratings of adviser's trustworthiness during the experiment. Participant's estimates of advice accuracy extracted from the overall winning model (3-level HGF, volatility tracking model) reflected the reported advisor fidelity on the randomly presented multiple choice questions (i.e., helpful, misleading, neutral). A linear regression analysis showed that participant's advisor ratings were predicted by their estimated advisor accuracy at the time of presentation of the multiple-choice question. The estimated beta parameter estimates were significantly different from zero ($t(85)=5.789$, $p<.001$). In addition, the wager magnitudes predicted by the model significantly correlated with the



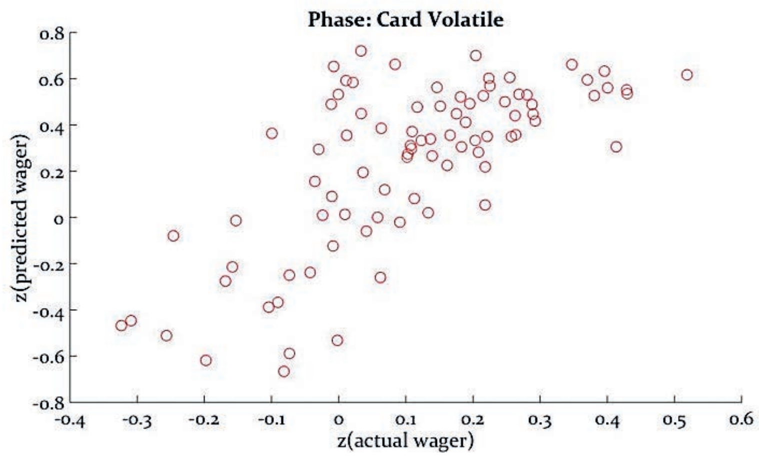


Figure 5.6 Model internal validity. The model predictions were consistent with participant's ratings of advisers fidelity across all four phases of the experiment.

participant's actual wagers. Across the four phases of the task, the predicted wager was highly correlated with the number of points participants actually wagered: advice stable *r*=.76, *p*<.001; advice volatile *r*=.77, *p*<.001; card stable *r*=.64, *p*<.001; card volatile *r*=.64, *p*<.001)(Figure 5.6).

5.3.1.2.3 Posterior Parameter Estimates

Three parameter that modulated the arbitration signal of the winning model are (1) κ or the coupling between the two hierarchical levels of the model that determined the impact of volatility on the inferred predictions of each information source, 2) ϑ that determines the variation of the volatility (meta-volatility), and 3) ζ, the social bias reflecting the reliance on the advice independent of its reliability. While there was no significant difference of ϑ across the two different learning domains, κ was significantly different (F(1,85)=42.91, *p*<.001, η_p^2=.34) (Figure 5.7). This result suggests that there is a difference in how participants learn from volatile reward probabilities and advisor fidelity.

Response model parameter ζ was not significantly different from one (*t*(85)=1.757, *p*=.083) (Figure 5.8). This indicates that, on average, participants did not prefer one information source over the other. κ_a was significantly associated with interpersonal traits (r=-.235, *p*=.029), which indicates that the information flow between χ2 and χ3 in the advice condition is reduced. This was not the case for the card condition. Furthermore, ϑ_a, ϑ_c as well as ζ were not associated with psychopathic traits.

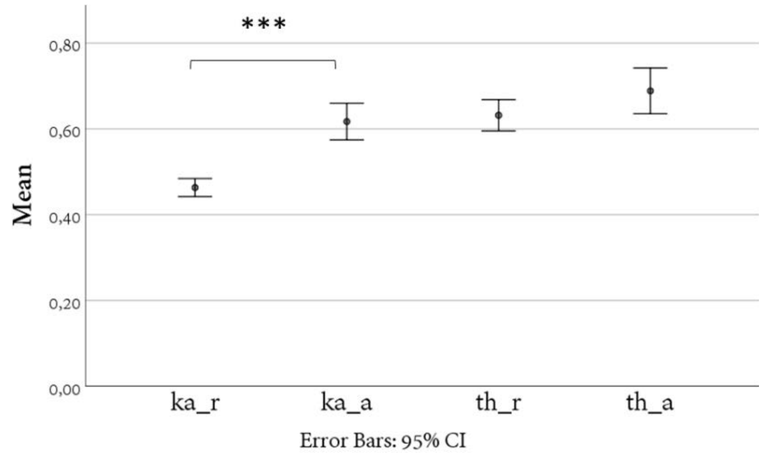


Figure 5.7 Learning model parameters. Results showed that there was a significant difference between κ for advice and κ for reward/card.

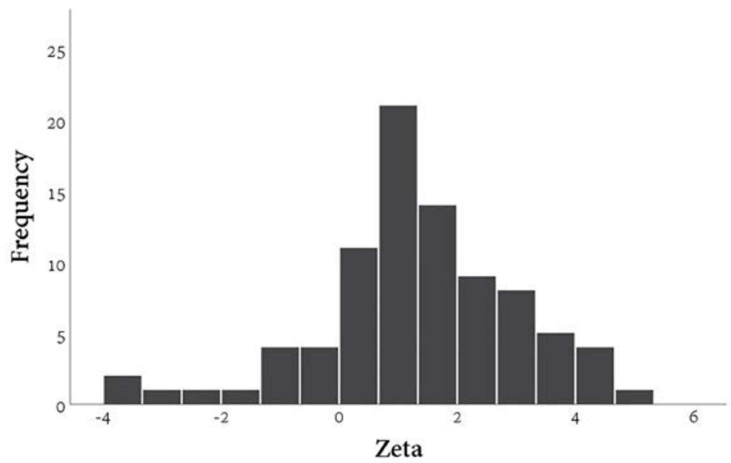


Figure 5.8 Response model parameter zeta (ζ). Distribution of zeta values.

5.3.2 EEG results

A topoplot and a time-frequency plot of the midfrontal electrodes show midfrontal theta power in Figure 5.9. We did not find main effects or an interaction effect for cue and phase. Performance accuracy was positively associated with theta power (*ρ*=.23, *p*=.039). More specifically, accuracy over volatile trials was positively associated with theta power on the corresponding trials (*ρ*=.29, *p*=.008). Accuracy over the other trial types was not associated with theta on those trials (Table 5.2). Furthermore, theta power over incorrect trials was higher as compared to correct trials (*t*(79)=-6.48, *p*<.001) (Figure 5.10).

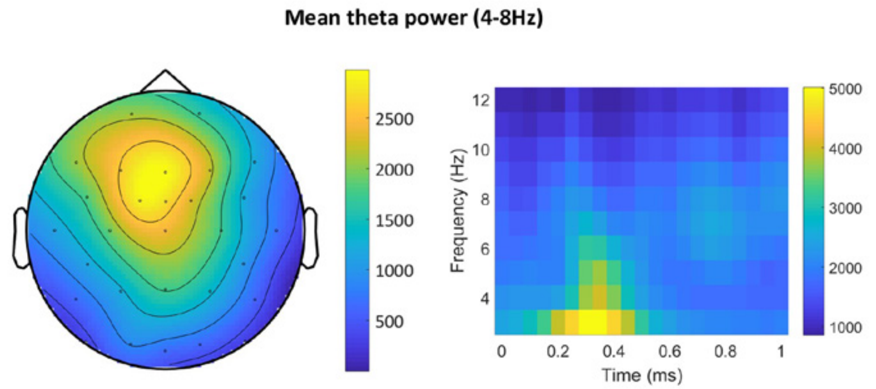


Figure 5.9 Mean theta power (4-8Hz). Left: Topoplot of mean theta power across 0.1-0.4ms after trial outcome representation (0ms). Right: Time-frequency plot of mean theta power (4-8Hz) over the frontal midline electrodes (Fz, FCz, Cz) across 0-1sec after trial outcome representation.

Analyses to investigate a possible association between theta power and the computational learning parameters showed us that ϑ for advice positively correlated with theta power (*ρ*=.22, *p*=.049). The other parameters did not show a significant correlation with theta power (Table 5.2; Bayesian correlations can be found in Table S5.2). Furthermore, we were interested in a possible link between psychopathic personality traits and mean theta power. Mean theta power over the midfrontal electrodes was negatively correlated with SRP total score (*ρ*=-.39, *p*<.001), as well as for the individual facet scores for interpersonal (*ρ*=-.35, *p*=.001), affective (*ρ*=-.37, *p*=.001), lifestyle (*ρ*=-.32, *p*=.004), and antisocial (*ρ*=-.36, *p*=.001) traits (Bayesian correlations can be found in Table S5.3).

Table 5.2 Spearman pairwise correlations between behavioral non-model and model parameters.

	Fz	FCz	Cz	midfrontal
	rho	*rho*	*rho*	*rho*
Performance accuray	.21	.27*	.23*	.23*
Accuracy volatile	.25*	.31**	.28*	.29**
Accuracy stable	.15	.21	.16	.16
Kappa card	-.04	-.02	-.07	-.03
Kappa advice	.10	.21	.23*	.20
Theta card	<.01	.04	<.01	.01
Theta advice	.12*	.22	.25*	.22*

Notes. Asterisks indicate significant correlations (*p<.05, **p<.01).

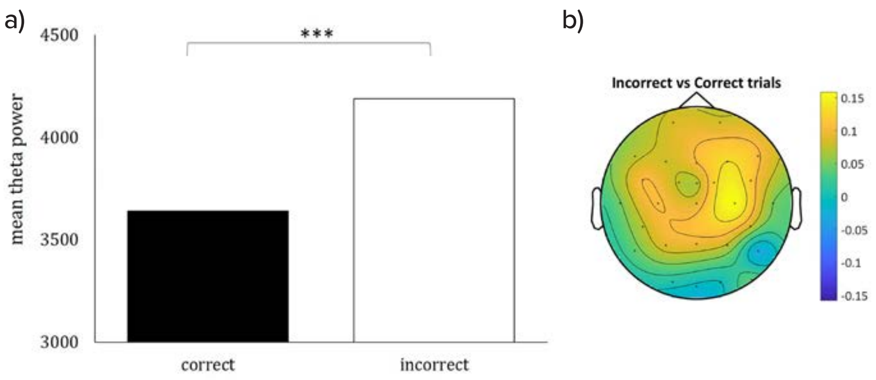


Figure 10. a) Mean theta power over the midline electrodes for correct and incorrect trials and **b)** Topographic representation of the TFR of the difference in theta power between incorrect and correct trials.

Table 5.3 Spearman pairwise correlations between SRP scores and mean theta power over the midline and the separate electrodes.

	Fz	FCz	Cz	midfrontal
	rho	*rho*	*rho*	*rho*
SRP Interpersonal	-.23*	-.30**	-.34**	-.35**
SRP Affective	-.24*	-.32**	-.39***	-.37**
SRP Lifestyle	-.27*	-.34**	-.32**	-.32**
SRP Antisocial	-.25*	-.36**	-.38***	-.36**
SRP Total	-.27*	-.36**	-.39***	-.39***

Notes. Asterisks indicate significant correlations (*p<.05, **p<.01, ***p<.001).

5.4 Discussion

The current study aimed to investigate how levels of psychopathic traits covary with i) the use social and non-social information during associative learning, ii) the updating of beliefs about others' trustworthiness, and iii) oscillatory theta band activation during associative learning. First, we will discuss the findings for the general, non-model, behavioral performance measures and their relationship with psychopathic traits. Then, we will consider the novel insights brought by the results obtained using the computational model and the electrophysiological measures, and how they may advance our understanding of psychopathy.

Our non-model behavioral findings showed that participants performed better and took social advice more into account during stable phases of the task as compared to volatile phases. Moreover, participants wagered more points during stable phases of the task, independent of whether the information was social or non-social. These findings demonstrate the effectiveness of the manipulation of volatility in our task. With regard to psychopathy, the results suggested that higher scores on erratic lifestyle and antisocial traits were associated with diminished performance during volatile phases of the task. This could be explained by considering the indications that excessive behavioral activation could be a key component underlying the erratic lifestyle and antisocial facets. Previous studies have proposed that individuals with high levels of lifestyle-antisocial psychopathic traits are characterized by a tendency to excessively pursue appetitive stimuli. This tendency has been associated with a decreased ability to properly monitor and adjust ongoing behavior in response to changes in environmental contingencies (Fowles, 1980; Newman, Patterson, & Kosson, 1987; Newman, Wallace, Schmitt, &

Arnett, 1997; Newman, MacCoon, Vaughn, & Sadeh, 2005; Buckholtz et al., 2010; Hoppenbrouwers, Neumann, Lewis, & Johansson, 2015). The association between volatility and psychopathic traits was significant on the social advice trials only, which is consistent with the proposed impairments in social learning and behavior in psychopathy (Blair, 2007; Brazil, Hunt, et al., 2013; Brazil et al., 2011). Finally, the amount of points wagered on each trial, reflecting their level of confidence about their choice, was not related to psychopathic traits. The latter result could be seen as an indication that these individuals still remained confident that they were making the right choice despite their relatively poor performance. Previous findings showed that monitoring of own behavior is not affected in psychopathic individuals (Brazil et al., 2011; 2009), which means that, in the light of prior findings, our result could not be attributed to poor performance monitoring. It does fit with prior reports suggesting an increased behavioral activation towards appetitive stimuli in individuals with high scores on the lifestyle-antisocial component of psychopathy. However, the present findings would argue that excessive tendency to pursue reward is not restricted to lifestyle-antisocial traits, but is related to the other facets as well.

One striking finding is that we did not find evidence for a relationship between psychopathy and the preference for using either reward or social advice (hypothesis 1), according to our non-model results that reflected the percentage of trials in which participants took the advice. This is in contrast with previous findings indicating that individuals with high levels of interpersonal-affective traits showed diminished use of reward during learning psychopathy (Blair, 2013; Brazil, Hunt, et al., 2013; Brazil et al., 2011), while individuals with high levels of impulsive-antisocial traits to hypersensitivity to reward showed increased use of reward-based information (Buckholtz et al., 2010). The overall performance accuracy was 60.1%, which indicated that participants experienced conflict on a relatively high number of error trials. This could have encouraged participants to switch between the two information sources, which made it harder to track their own performance accuracy on each of the two information sources.

While informative, these results do not elucidate how psychopathy might be related to disturbances in the individual cognitive processes that operate within the associative learning mechanism. Therefore, a computational model was used to investigate the latent cognitive computations that underlie the pattern behavioral performance of each participant. First, the results from the model comparison indicated that participants learned according to a mechanism in which they tracked the rate of change and the overall volatility of both social and non-social information, which is in line with prior research (e.g. Diaconescu et al., 2014, 2017). Analyses concerning the computational parameters demonstrated a difference in the kappa parameter, which quantifies the flow of information between the two hierarchical levels that quantify the advisor's trustworthiness and the current rate of change of the advisor's intentions. This result could indicate that it may be more difficult to estimate the likelihood and the rate of change of social advice relative to reward in general. During reward-based learning, the individual can rely on outcome information that is relatively easy to identify (e.g. a coin indicating reward), while a social context introduces additional levels of complexity that requires the involvement of higher-order cognitive processes in order to determine the valence of the outcomes. We suggest that stronger coupling between the two levels of volatility in a social context results from an increased sense of uncertainty about the trustworthiness of social information. That means, the use of reward-based information requires us to rely on our own estimates, while using social information requires us to rely on the trustworthiness of the advisor. We experience more control during the former as compared to the latter.

With regard to psychopathic traits, we found that the interaction between the estimated volatility of the trustworthiness of the advisor and the estimated tendency of the advisor to give accurate advice was reduced (i.e. kappa values were lower) with increasing levels of interpersonal traits. One possible explanation could be that individuals with high interpersonal traits do not sufficiently process the complexity of the social context. Therefore, it could be that some cognitive processes required for proper evaluation of the social context are less engaged (see Brazil ea 2011, 2013), resulting in a smaller need to update beliefs concerning the volatility of social advice. The lower values of kappa could reflect this relative 'disengagement' of the mechanism in this regard. This explanation aligns with our other behavioral findings indicating a negative relation between psychopathic traits and performance accuracy, which were most prominent on the volatile advice trials. Furthermore, it is consistent with the extensive amount of evidence indicating impairments in associative and social learning in psychopathy (Blair, 2007; Blair et al., 2004; Brazil, Hunt, et al., 2013; Brazil, Maes, et al., 2013; Gregory et al., 2015; Mitchell et al., 2006; Von Borries et al., 2010). Other model-based findings showed that the perceived trustworthiness of the advice was equally stable across information sources (social vs. reward-based), and was not related to psychopathic traits. Moreover, in line with our non-model findings, the response model parameter reflecting the social bias showed that participants did not prefer one information source over the other, and this was not influenced by the level of psychopathic traits. Taken together, our behavioral results support the hypothesis that individuals with psychopathic traits experience impairments in associative learning based on social information, and suggest that interpersonal traits are linked to a reduced ability to adapt to changes in the reliability of social information. According to our results, this did not lead to a preference for one of the information sources and did not affect the risk that was taken in order to obtain a high reward.

The increased precision offered by the modelling approach was further enhanced by analysis of electrophysiological measurements of the mPFC, as direct measures of adaptive control during associative learning. As midfrontal theta power has been suggested to play an important role in the implementation of adaptive control in associative learning, we recorded theta band EEG activity over the midfrontal electrodes during task performance and investigated a possible link with psychopathic traits. Our results showed no difference in theta power during volatile as compared to stable trials. We did find that performance accuracy was positively associated with theta power. This indicates that increased theta activation is associated with higher performance, which is in line with previous findings indicating that event-related theta activity is positively related to cognitive performance (Klimesh, 1999). Moreover, the absence of a difference in theta during stable as compared to volatile phases suggests that theta power is not associated with volatility monitoring. This is in contrast with our other findings suggesting that the parameter that represents the variation of the volatility (meta-volatility) was positively related to theta power. The latter finding supports the evidence for a role for theta power in the implementation of adaptive control in associative learning that was proposed previously (Behrens et al. 2007). Future studies including community samples are needed to examine this effect in more detail. Finally, we found that overall, theta activation was negatively related to psychopathic traits (total and facet scores). This is in line with our findings indicating a negative relationship between performance accuracy and psychopathic traits, and a negative relationship between theta power and performance accuracy. Meta-analytical findings showed that theta plays a key role in the processing of emotion (Knyazev, 2007) and increased theta synchronization was found in individuals scoring high on measures of emotional intelligence (Knyazev et al., 2013). The lower theta power that we found in relation to psychopathic traits are consistent with the emotional deficits and decreased emotional intelligence that were found in psychopathic individuals (Megías, Gómez- Leal, Gutiérrez-Gobo, & Fernández-Berrocal, 2018).

The findings of the present study highlighted impairments in associative learning and provided further insight into the individual cognitive processes that operate within the associative learning mechanism. In addition, it examined the role of theta activation in these learning processes. Our results indicated that increased psychopathic traits were i) not associated with a preference for social or non-social information in the Wager task, ii) associated with impaired belief updating about other's trustworthiness, and iii) associated with decreased theta power during associative learning. Taken together, our findings indicate that individuals with psychopathic traits experience impairments in associative learning of social information and suggest that interpersonal traits are linked to a reduced ability to adapt to changes in the reliability of social information. Decreased theta power has been found to be linked to higher levels of psychopathic traits, and there are indications that theta is involved in tracking the volatility of social information. While prior studies highlighted the link between theta and cognitive performance, and theta and psychopathic traits separately, this is the first study that provided support for the triangular connection between the three concepts. Finally, we found that the impairments that were associated to increased levels of psychopathic traits, mostly present in the social domain, did not lead to a preference for one of the information sources, and did not affect the risk that was taken in order to obtain a high reward. This information could attribute to our understanding about the underlying associative learning mechanisms that cause low responsivity towards current treatment interventions in those with psychopathy.

5.5 Supplemental material

Table S5.1 Bayesian pairwise correlations between SRP scores and performance on volatile advice trials

	Volatile Advice trials	
	Mode r	**95% CI**
SRP Interpersonal	-.28*	[-.47, -.07]
SRP Affective	-.31*	[-.49, -.10]
SRP Lifestyle	-.22*	[-.42, -.01]
SRP Antisocial	-.18	[-.38, .04]
SRP Total	-.31*	[-.49, -.09]

Notes. Asterisks indicate significant correlations (*p<.05).

Table S5.2 Bayesian pairwise correlations between behavioral and non-model and model parameters

	Fz		FCz		Cz		midfrontal	
	Mode r	95% CI	Mode r	95% CI	Mode r	95% CI	Mode r	95% CI
Performance accuray	.18	[-.09, .38]	.24*	[.03, .43]	.23*	[.01,.42]	.22*	[.00, .41]
Acc VA trials	.21	[-.01, .40]	.25*	[.03, .44]	.29*	[.08, .47]	.25*	[.03, .44]
Kappa card	-.09	[-.30, .13]	-.06	[-.27, .16]	-.11	[-.31, .11]	-.09	[-.29, .13]
Kappa advice	.16	[-.06, .36]	.22*	[.01, .42]	.22*	[.00, .41]	.20	[-.01, .40]
Theta card	.02	[-.20, .23]	.08	[-.14, .29]	.07	[-.14, .28]	.06	[-.16, .27]
Theta advice	.20	[-.02, .40]	.23	[-.02, .43]	.23*	[.01, .42]	.22*	[.01, .42]

Notes. Asterisks indicate significant correlations (*p*<.05). Acc VA trials = accuracy volatile advice trials.

Table S5.3 Bayesian pairwise correlations between SRP scores and mean theta power over the midfrontal and the separate electrodes

	Fz		FCz		Cz		midfrontal	
	Mode r	95% CI	Mode r	95% CI	Mode r	95% CI	Mode r	95% CI
SRP Interpersonal	-.28*	[-.47, -.07]	-.34*	[-.52, -.13]	.34*	[-.52, -.14]	-.33*	[-.50, -.17]
SRP Affective	-.25*	[-.44, -.03]	-.34*	[-.52, -.13]	-.38*	[-.55, -.17]	.32*	[-.50, -.12]
SRP Lifestyle	-.18	[-.38, .04]	-.25*	[-.44, -.04]	-.28*	[-.46, -.06]	-.24*	[-.43, -.02]
SRP Antisocial	-.11	[-.32, .10]	-.19	[-.39, .02]	-.29*	[-.43, -.03]	-.18	[-.38, .04]
SRP Total	-.26*	[-.45, -.05]	-.35*	[-.53, -.14]	-.38*	[-.55, -.18]	-.33*	[-.51, -.13]

Notes. Asterisks indicate significant correlations (*p*<.05).

6

Moral strategies and psychopathic traits

This chapter is based on:
Driessen, J. M. A., van Baar, J. M., Sanfey, A. G., Glennon, J. C., & Brazil, I. A. (in press). Moral strategies and psychopathic traits. *Journal of Abnormal Psychology.*

Abstract

Individuals with elevated psychopathic traits often make decisions that have a negative impact on others. Some findings suggest that a lack of empathy and guilt is a key explanatory factor, while other results point towards a decreased sense of fairness in individuals with elevated psychopathic traits. The goal of the present study was to directly compare these hypotheses. Eighty-six healthy individuals completed the Self-Report Psychopathy scale and performed the Hidden Multiplier Trust Game, a socio-economic decision-making task designed to untangle the roles of guilt and fairness during decision-making. Computational modelling of choice data identified five types of moral decision strategies: inequity aversion, guilt aversion, moral opportunism, greed, and generosity. The results indicated that the affective traits that are associated with psychopathy were linked to lower levels of reciprocity, which indirectly reflects greed. Our computational analyses provided more insight into the strategies that underlie reduced reciprocity behavior in individuals with a high level of affective traits. They indicated that a reduced sense of fairness, but not guiltlessness underlie the reduced reciprocity behavior in individuals with high levels of affective traits. This could explain the link between affective traits and poor social decision-making and immoral behavior in individuals with elevated psychopathic traits. Our findings stress the importance of treating guilt and fairness as independent concepts, and it is possible that the lack of methodological and conceptual precision in untangling the individual impact of fairness and guilt in previous studies could explain the mixed results in moral decision-making literature. Elucidating the psychological motivations underlying the relationship between psychopathic traits and poor moral decision-making opens new avenues for research on the underlying cognitive mechanisms.

6.1 Introduction

Prosocial behavior can be defined as taking voluntary actions intended to benefit one or more persons other than oneself, and involves behaviors such as helping, sharing, and cooperating (Millon, Lerner, & Weiner, 2003). Prosocial behavior is a key element in daily life, and much of our joy —and suffering— is dependent on the pro- and antisocial choices made by other people. While most people behave prosocially in general, there are individuals who systematically engage in severely disruptive antisocial behaviors. These individuals often have elevated levels of psychopathy, a personality construct that encompasses a constellation of traits reflecting emotional dysfunction and antisocial behavior. These traits concern interpersonal functioning (e.g. manipulativeness, grandiosity), affective disruptions (e.g. lack of empathy, guiltlessness), the inclination to lead an erratic lifestyle (e.g. irresponsible behavior, sensation-seeking) and the engagement in antisocial acts (e.g. criminality) (Hare, 2003b). Individuals with the most severe levels of psychopathic traits are often seen in penitentiaries, but psychopathic traits can also be found and measured dimensionally in the general community (DeMatteo et al., 2006; Neumann & Hare, 2008). Community-dwelling individuals with elevated psychopathic traits tend to make choices that can cause great suffering to their families, friends, and colleagues (Mathieu & Babiak, 2016), and research has repeatedly linked psychopathy to poor social decision-making and pervasive immoral behavior (Blair, 1995; Blair, 2013).

Importantly, an understanding as to why individuals with psychopathic tendencies make bad choices requires much better knowledge of the general cognitive mechanisms that could be involved in social decision-making (Brazil et al., 2013). One prominent line of thinking suggests that prosocial choices are driven by fairness considerations, i.e. that a distaste for unfair outcomes motivates prosocial behavior in humans. The perception of (un)fairness can be viewed as the degree to which an individual dislikes unequal outcomes and is formalized in computational models of inequity aversion (Bolton & Ockenfels, 2000; Brosnan, 2006; Fehr & Schmidt, 1999). Inequity-averse individuals are often willing to sacrifice some of their own payoff in order to ensure more equitable outcomes with others. Research in this domain has primarily employed socio-economic decision-making paradigms in order to investigate people's feeling of fairness regarding their own material payoff relative to the payoff of others. Previous findings resulting from these studies suggest that, in addition to their own payoff, people value the equity in outcome between themselves and others (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). Studies on the role of fairness considerations during social decision-making in psychopathy have yielded mixed results. Some findings point towards a reduced sense of fairness (Aharoni, Antonenko, & Kiehl,

6

2011; Glenn, Iyer, Graham, Koleva, & Haidt, 2009; Osumi & Ohira, 2010). For instance, Osumi and Ohira (2010) found that individuals with elevated levels of psychopathic traits, in contrast to individuals with low levels of psychopathic traits, accepted more unfair offers and did not show a fairness-dependent modulation of skin conductance response in an Ultimatum Game. In another study in non-offenders, participants with elevated psychopathic traits rejected a similar amount of unfair offers relative to individuals with low psychopathic traits, but perceived unfair offers as being less unfair (Vieira et al., 2014). In contrast, Koenigs and Tranel (2007) found the opposite effect in psychopathic offenders, showing that offenders with high levels of interpersonal-affective traits had lower acceptance rates to unfair offers and showed a reduced amount of reciprocity. This was further supported by studies reporting an absence of a direct association between perceived fairness and acceptance rates in individuals with high psychopathic tendencies (Radke et al., 2015; Vieira et al., 2014). Together, the mixed results highlight that the association between sense of fairness and psychopathy is still unclear.

An alternative theory proposes that guilt aversion is the main driver of prosocial behavior. Guilt can be conceptualized as a negative emotional state that occurs when one believes to have inflicted harm, loss, or distress to another person (Baumeister, Stillwell, & Heatherton, 1994). We typically generate beliefs about what we think others expect us to do (i.e., second-order beliefs), and the feeling of guilt can emerge as a result of not living up to these expectations. Accordingly, we are motivated to cooperate, not because of prosocial feelings, but rather because we anticipate feeling bad (i.e. guilty) if we fail to live up to the expectations of others and will try to avoid this (Baumeister et al., 1994). Prior studies demonstrated that people are indeed guilt-averse and, in fact, often do make decisions to minimize their anticipated guilt regarding a social interaction (Chang, Smith, Dufwenberg, & Sanfey, 2011; Charness & Dufwenberg, 2006; Dufwenberg & Gneezy, 2000; Khalmetski, 2016). Thus, this type of social guilt functions as a feedback mechanism on either executed or imagined behavior and the anticipation of feeling guilt prevents us from engaging in antisocial or immoral behavior (Prado, Treeby, & Crowe, 2016; Seara-Cardoso, Dolberg, Neumann, Roiser, & Viding, 2013). There are findings highlighting that an aberrant experience of guilt subserves social decision-making in psychopathy (Blair, Peschardt, Budhani, Mitchell, & Pine, 2006; Cleckley, 1964; Seara-Cardoso, Sebastian, Viding, & Roiser, 2016). For instance, participants' reported degree of guilt aversion in a social monetary reward game was negatively associated with interpersonal-affective psychopathic traits (Gong, Brazil, Chang, & Sanfey, 2019). On the neural level, higher levels of psychopathic traits have been linked to reduced modulation of activity in the insula in response to guilt (Seara-Cardoso, Neumann, Roiser, McCrory, & Viding, 2012).

Research on the employment of strategies reflecting inequity aversion and guilt aversion has been important for our current understanding of social decision-making, with ample evidence supporting their explanatory power. However, recent work has highlighted three alternative moral decision strategies. First, some individuals are neither inequity- nor guilt-averse, and instead make decisions mainly driven by greed. These individuals seek to maximize personal gains, even if it is at the expense of another person. Second, some individual's choices are characterized by moral opportunism, which reflects the tendency to employ a strategy that is both morally justifiable and maximally financially lucrative, depending on the situation (van Baar, Chang, & Sanfey, 2019). Moral opportunists thus switch between different modes of moral reasoning (inequity aversion and guilt aversion), which has been corroborated by neuroimaging work in which moral opportunists alternate between neural activity patterns associated with inequity aversion and guilt aversion (van Baar et al., 2019). Third, research on altruism also points towards the existence of a fifth strategy which we will refer to as generosity. Generosity can be seen as the avoidance of any inequity and guilt, and this type of altruism was proposed to be the exact opposite of psychopathic behavior (Marsh et al., 2014).

The current understanding of the relationship between psychopathic traits and the different strategies underlying social decision-making is limited for several reasons. First, prior studies have yielded contradicting results on the link between psychopathy and inequity aversion or guilt aversion. Second, prior work did not consider all of the socio-economic decision-making strategies outlined above. Third, a key disadvantage of many laboratory paradigms is that different moral strategies yield the same behavioral outcomes. That is, in common tasks like the Ultimatum Game or Trust Game, inequity-averse behavior and guilt-averse behavior are indistinguishable (Hein, Morishima, Leiberg, Sul, & Fehr, 2016; Nihonsugi, Ihara, & Haruno, 2015; van Baar et al., 2019).

Fortunately, recent methodological advancements make it possible to precisely determine the moral decision strategy of participants by combining new tasks with computational modeling (Gong et al., 2019; van Baar et al., 2019), which provides valuable insight into the motivational bases of social decisions. In the present study, we use this approach to disentangle the moral decision strategies that reciprocity behavior and relate these to psychopathic traits. Our computational model yields motive-specific parameters that allows us to index the strength of a participant's inclination towards employing a particular moral strategy such as guilt aversion (GA) and inequity aversion (IA). Given that guiltlessness is believed to be a core feature of psychopathy (Gong et al., 2019; Hare, 2003; Seara-Cardoso et al., 2016; van Baar et al., 2019), we hypothesized that the affective facet of psychopathy should be negatively associated with the computational parameter representing guilt aversion. The mixed literature on fairness concerns in psychopathy make it difficult to generate clear predictions for the parameter representing inequity aversion. Previous literature indicated a hypersensitivity to reward (Baskin-Sommers, Wallace, MacCoon, Curtin, & Newman, 2010; Buckholtz et al., 2010) and selfish behavior (Mokros et al., 2008; Rilling et al., 2007) in individuals with psychopathic traits. Therefore, a relation between greed and psychopathic traits was expected to be reflected by a negative association with reciprocity and a high presence of these traits. We refrained from generating clear hypotheses for moral opportunism and generosity given that these strategies have not been explored before in relation to psychopathy.

6.2 Methods

6.2.1 Participants

Based on a priori samples size calculations[2], eighty-seven participants (36 males) were included in the study. They were between 18 and 35 years of age (*M*=24.1, *SD*=2.6), did not suffer from a neurological or psychiatric disease, and were not using any psychoactive substances. These participants were selected from a large database (n=1519, 309 males) based on their level of psychopathic traits, whom were recruited via local advertisement and an online research participation system (SONA systems) of the Radboud University in Nijmegen, The Netherlands. We employed an oversampling procedure in order to include more individuals with psychopathy scores falling in the tails of the distribution in the population, which would otherwise remain under-represented (Hall, Bernat, & Patrick, 2007; Bernat, Nelson, Steele, Gehring, & Patrick, 2011; Brazil et al. 2013; Gong, Brazil, Chang, & Sanfey, 2019). The total scores in the large sample (*N*=1519, 309 males) were divided in quartiles to make sure that 25% of the participants ended up in the top and bottom quartiles, while 50% of the participants ended up in the two middle quartiles. We selected 87 participants (36 males) based on their SRP total score for the current study. The top and bottom quartiles were oversampled in order to enhance the presence of extreme scores on both sides of the distribution (8, 60). Consequently, 20 participants (23.3%) belonged to the lowest quartile, 40 participants (46.5%) belonged to the two middle segments, and 26 participants (30.2%) belonged to the upper quartile.

Participants' level of psychopathic traits were assessed with the Dutch version of the Self-Report Psychopathy checklist–Short Form (SRP-SF, Dutch version; Gordts et al., 2017), a well-validated self-report questionnaire consisting of 29 items that need to be rated on a 5-point Likert scale (Gordts et al., 2017; Hare & Neumann, 2008; Lilienfeld et al., 2006; Neumann & Pardini, 2014). It yields scores on four subscales (interpersonal, affective, lifestyle, and antisocial) and a total score (Table 6.1). Internal consistency in our sample was high (all Cronbach's α >.75).

2 Sample size calculations based on previous work using a computational approach to study social decision-making in relation to psychopathy (Brazil et al., 2013) indicated that approximately 88 participants were required for the project (power=0.85).

Table 6.1 Means and SDs of the SRP scores of each group used in our oversampling procedure and the total sample.

	Low (n=20)		Intermediate (n=40)		High (n=26)		Total (n=86)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SRP Interpersonal	8.00[a]	1.12	10.85[b]	2.46	18.15[c]	4.29	12.40	4.93
SRP Affective	8.15[a]	1.18	10.25[b]	2.15	15.89[c]	3.49	11.45	3.91
SRP Lifestyle	9.00[a]	1.52	12.42[b]	2.90	17.15[c]	3.77	13.06	4.21
SRP Antisocial	7.00[a]	0.00	7.10[a]	0.38	9.04[b]	3.39	7.66	2.07
SRP Total	32.15[a]	2.03	40.60[b]	4.24	60.23[c]	10.53	44.56	12.67

Notes. Means with different superscripts differ significantly from each other (*p*<.05).

One of the male participants did not adhere to the task instructions and was excluded from the analyses, therefore the final number of participants included in the analysis is eighty-six. Participants provided written informed consent prior to the onset of the study, and received monetary compensation at the end of the study. All procedures were approved by the Social Science Ethics Committee of the Radboud University Nijmegen, Netherlands (ECSW2017-0805-512) and was carried out in accordance with the standards of the declaration of Helsinki.

6.2.2 Hidden Multiplier Trust Game

The Hidden Multiplier Trust Game (HMTG) (Figure 6.1) is a modified version of a traditional Trust Game (van Baar et al., 2019). On each trial, an anonymous Investor can invest any number of 10 coins in a Trustee, retaining the remainder. The Investor believes that his investment is multiplied by a fixed factor before it is transferred to the Trustee. The Trustee can then choose to return any number of the transferred coins to the Investor, but does not have to do so. In this version of Trust Game, the Investor is told that the fixed multiplier is ×4. Importantly, an in contrast to a typical Trust Game, the Trustee is aware that the actual multiplier is not fixed, but can in fact be either ×2 (25% of the trials), ×4 (50% of the trials) or ×6

(25% of the trials). Moreover, the Trustee is aware of the Investor's ignorance as to the hidden multiplier, and thus knows that the Investor believes that the multiplier is always ×4. This results in an information asymmetry between the two players, and therefore different moral strategies predict different Trustee decisions when the multiplier is different from ×4. An inequity-averse Trustee, who aims to ensure an even split, will decide based on the actual amount of coins he receives, which in turn depends on the true multiplier used on any given trial. A guilt-averse Trustee, who is eager to meet the Investor's expectations in order to avoid feelings of guilt, should always return the amount of coins that were expected based on the Investor's belief of a fixed ×4 multiplier. His decisions are therefore made irrespective of the actual multiplier on that trial. A greedy Trustee keeps as many coins as possible for himself. A morally opportunistic Trustee displays inequity-averse behavior in the ×2 condition, but guilt-averse behavior in the ×6 condition (van Baar et al., 2019). This participant always follows a non-greedy moral strategy

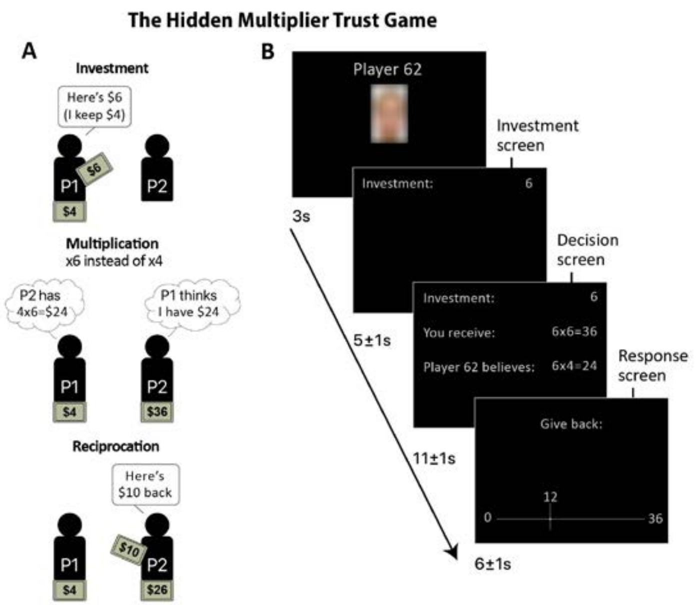


Figure 6.1 Panel A: Hidden Multiplier Trust Game. Schematic representation of the task with Player 1 (P1) as Investor and Player 2 (P2) as Trustee. Investment: P1 invests 0-10 coins. Multiplication: The investment of P1 is multiplied by ×2, ×4, or ×6. Reciprocation: P2 decides the amount of coins to return to P1. The participants in the present study always played the role of the Trustee. Panel B: Representation of the trial timeline. Reproduced from Van Baar et al. (2019).

that is nevertheless maximally financially lucrative within the constraints of prosocial behavior. A generous Trustee ensures that guilt and inequity are always minimized. This strategy predicts that a Trustee is guilt-averse in the ×2 and ×4 condition, but inequity-averse in the ×6 condition. This results in a generous behavioral pattern in which the Trustee aims to meet the Investor's expectations and offers even a little more when the multiplier is higher than expected.

6.2.2 Procedure

The present study is part of a larger project in which participants played three different computer tasks. The order of the tasks was randomized between participants and tasks other than the HMTG are not considered in this chapter. Participants received instructions before the start of the HMTG, and the experimenter checked whether each participant understood the instructions correctly. In order to avoid possible biases caused by the information provided, the HMTG was referred to as the 'Investment game', the Investor as 'Player 1' and the Trustee as 'Player 2'. Each participant was instructed that he/she would play the role of Player 2 against 90 anonymous Players 1. Each Player 1 was assigned a unique number and a blurred photo of a face to make clear that each round was played against a different Player 1. The photos were taken from the Radboud Faces Database (Langner et al., 2010) and blurred heavily using MATLAB (v2013a, Mathworks, Natick, MA, USA), which made it impossible to infer the gender. Participants were also told that each Player 1 had previously participated in a study that used this paradigm and had agreed to letting us use his/her data in the following studies. All stimuli were presented using PsychToolBox 3.0.11 (www.psychtoolbox.org) in MATLAB. To increase ecological validity and participants' motivation, participants were told that they would receive a bonus that was determined by one randomly selected trial at the end of the experiment. They were told that this trial would be financially consequential for Player 1 as well. The coins earned in this trial were redeemed for actual money (1 coin=10 cents). Participants were asked to make their own investment as Player 1 to enhance plausibility of the task. They were told that this could potentially be used in future studies, and that in case their trial was selected for a bonus payment, they would be contacted via the participation system and would receive the payment afterwards.

6.2.4 Analysis

6.2.4.1 Reciprocity behavior

The mean percentage of the amount of coins returned (i.e. reciprocity behavior) was computed across all trials as well as per multiplier condition. In order to examine the link between reciprocity and psychopathic traits, we first performed

Bayesian pairwise correlations between the reciprocity conditions (all trials, ×2, ×4, and ×6, respectively) and the total score and facet scores of the SRP-SF. Mode r, representing the most likely estimate of the strength of the correlation, and the corresponding 95% credibility intervals are reported.

A linear mixed effects regression was performed in R (version 1.1.456, R Core Team, 2018) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2014) in order to investigate the relationship between reciprocity and the investment (0-10 coins) and multiplier (×2, ×4, ×6) and to evaluate the potential of the HMTG to elicited reciprocity behavior. We started with the maximal linear mixed effect model justified by the structure of the data. Investment (continuous predictor), multiplier (factor with 3 levels) and their interaction term were entered as fixed effects. A random intercept for subjects and by-subject random slopes of investment and multiplier were also included. We simplified the model as needed due to non-convergence and removed the random slopes that were near zero in order to avoid overfitting. Reported p-values were obtained using the Sattherwaite approximation to degrees of freedom in the lmerTest (Christensen, 2019).

6.2.4.2 Moral strategy model

To measure the moral decision strategy employed by each participant in the HMTG, we fit a utility model, the Moral Strategy Model (MSM), to the observed choice behavior in this task. This model was inspired by the model used by van Baar and colleagues (2019), with an adjustment to the functional form to incorporate the fifth predicted moral strategy of generosity within the model's space of predicted choice patterns. The MSM posits that the utility derived by the Trustee from their reciprocity decisions (U_2) is jointly determined by the financial payoff earned by the Trustee (π_2), the inequity created by the decision, and the Trustee's aversion towards feeling guilt. In the model, the relative influence of payoff, inequity, and guilt on overall utility is controlled by two free parameters phi (φ) and zeta (ζ), which place a weight on guilt and inequity, respectively:

$$U_2 = \pi_2 - \varphi \times Guilt_2 - \zeta \times Inequity_2 \quad (1)$$

Trustee payoff is defined as $\pi_2 = (I \times M_2 - S_2) / (I \times M_2)$, where I is the Investor's investment amount, M_2 is the multiplier known only to the Trustee, and S_2 describes the Trustee's strategy (i.e., the amount of money to return in the game); IA is based on previous formulations with $Inequity_2 = ((I \times M_2 - S_2) / (10 - I + I \times M_2) - ½)^2$ (Bolton & Ockenfels, 2000; van Baar et al., 2019); as is GA with $Guilt_2 = max(((E_2(E_1(S_2)) - S_2) / (E_1(M_1) \times I)), 0)^2$ (Battigalli & Dufwenberg, 2007; Dufwenberg & Gneezy, 2000; van Baar et al., 2019), where $E_2(E_1(S_2))$ refers to the Trustee's second-order belief about the Investor's expectations of the Trustee's strategy and $E_1(M_1)$ refers to the

Investor's belief about the multiplier (always ×4). The inequity and guilt models both contain a quadratic term, ensuring that increasing inequity or guilt has an outsized influence on behavior. This nonlinearity is required to predict subtle changes in behavior (Fehr & Schmidt, 1999; van Baar et al., 2019). However, the guilt model also contains a maximum term max(), which ensures that cases in which the Trustee returns more coins (S_2) to the Investor than expected ($E_2(E_1(S_2))$, guilt is set to zero, as this would not constitute disappointment of the Investor.

We fit this model to individual participant data by varying phi and zeta between the parameter bounds (0 and 1) and minimizing the sum of squared model error using the least_squares function from the *Scipy* package for Python. We initialized this fitting procedure 100 times per participant, each time starting the algorithm at a random point in phi-zeta parameter space, to avoid ending the procedure in a local maximum. This model makes it possible to draw inferences about the moral strategy of the participant. For example, a participant whose behavior is best

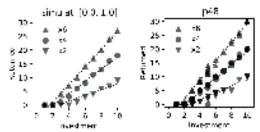
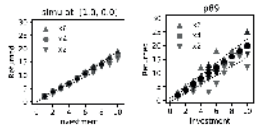
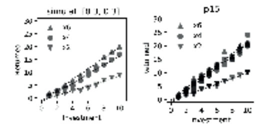
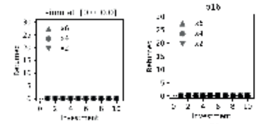
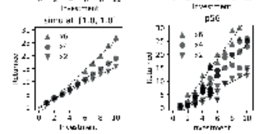
Moral strategy	Description	Parameter values [phi, zeta]	Behavioral pattern / Participant example
Inequity aversion (IA)	Aims to ensure an even split	[0,1]	
Guilt aversion (GA)	Aims to meet the Investor's expectations in order to avoid guilt feelings	[1,0]	
Moral opportunism (MO)	Always follows a non-greedy moral strategy that is nevertheless maximally financially lucrative within the constraints of prosocial behavior	[0.3, 0.3]	
Greed (GR)	Keeps as many coins as possible for himself	[0,0]	
Generosity (GE)	Ensures that guilt and inequity are always minimized	[1,1]	

Figure 6.2 Overview moral strategies in the Hidden Multiplier Trust Game. Column 1: Label, Column 2: Description, Column 3: Parameter values [phi, zeta], Column 4: Simulated behavioral response patterns, Column 5: Observed behavioral patterns of five example participants. Grey symbols reflect the returned amount of coins on each trial (black symbols result from an overlay of two or multiple grey symbols).

captured by the model with high phi and low zeta can be said to be guilt-averse in their behavior. In the same way, all four previously validated moral strategies for the HMTG (van Baar et al., 2019)—IA, GA, moral opportunism (MO), and greed—can be captured by different pairs of free parameters: GA at parameter values near φ=1 and ζ=0; IA near φ=0 and ζ=1; MO near φ=0.3 and ζ=0.3; Greed near φ=0 and ζ=0. In addition, the model captured a fifth strategy which we labeled generosity, with parameter values near φ=1 and ζ=1 (Figure 6.2, 3rd column).

In order to provide a categorical clustering of participants by their decision strategy, we clustered the two-dimensional phi-zeta parameter space of the model into five zones where the five moral strategies are best represented. We first drew 'moral strategy exemplar' simulations (Figure 6.2, 4th column) of the model at the five aforementioned parameter pairs (e.g. φ=0 and ζ=1 for IA), then ran 10201 simulations of the model at 101x101 equidistant parameter pairs between in ranges 0<=φ<=1 and 0<=ζ<=1, and labeled each of these 10201 parameter pairs by the exemplar simulation to which it was most similar based on the sum of squared error. For instance, for the model simulation at φ=0.15 and ζ=0.70, we computed the sum of squared difference with each of the five exemplar simulations of the five moral strategies, found that it was most similar to the exemplar simulation of guilt aversion, and labeled it 'GA' for guilt aversion. We then used the model as fitted to each participant's decision data to place that participant in the parameter space, and labeled the participant by the moral strategy zone in which they were placed.

6.2.4.3 Model performance

In order to investigate the performance of our model, we compared the model fit of our model to that of the component models of greed (eq. 2), inequity aversion (eq. 3), and guilt aversion (eq. 4), where payoff, inequity, and guilt were defined identically to these terms in the moral strategy model:

$$U_2 = \pi_2 \tag{2}$$
$$U_2 = \pi_2 - \zeta \times \text{Inequity}_2 \tag{3}$$
$$U_2 = \pi_2 - \varphi \times \text{Guilt}_2 \tag{4}$$

Additionally, we conducted parameter recovery tests on the model to ensure that our model was not overfitting or unreliably fitting the data. In order to do so, 86 subjects were simulated by simulating behavioural data at 86 random points in the model parameter space. Our model was fitted to these simulated subjects using the same fitting procedure as for the participant data. The correspondence between the true and recovered parameters as well as the correspondence between the true simulated task data and the task data predicted based on the recovered parameters were tested.

6.2.4.4 Behavioral parameters

Bayesian pairwise correlations were performed to assess the associations between the participant performance (percentage of coins returned) and the computational parameters as well as the psychopathy factor and total scores. Furthermore, bayesian pairwise correlations were performed to test the associations between psychopathic traits and the moral strategies by using the model parameters phi, zeta, the SRP-SF subscale and the total scores. Bayesian correlations do not rely on any fixed assumptions about the distribution of the data (and is better at handling data that would be considered problematic in traditional approaches), do not suffer from many other empirical issues that plague traditional parametric and non-parametric frameworks (Morey & Wagenmakers, 2014), and is a common approach in computational neuroscience.

6.3 Results

6.3.1 Reciprocity behavior

The mean percentage of the amount of coins that was returned by the Trustees was 33.37% (*SD*=0.11). The average amount of coins returned for each of the multiplier separately were as follows; ×2 = 31.26% (*SD*=0.14), ×4=34.71% (*SD*=0.11), and ×6=32.66% (*SD*=0.11). A comparison of the reciprocity outcomes between the moral strategies over all three conditions can be found in Table 6.2.

Table 6.2 Percentage returned amount of coins for each condition

	IA	GA	MO	GR	GE
all trials	.37*	.45*	.33	.02*	.49*
x2 trials	.31	.66*	.32	.01*	.60*
x4 trials	.36	.47	.36	.02*	.47*
x6 trials	.40*	.37	.30	.03*	.48*

Notes. Asterisks (*) indicate a significant difference with MO (p<.05). IA=inequity aversion, GA=guilt aversion, MO=moral opportunism, GR=greed, GE=generosity.

We examined the association between reciprocity and psychopathic traits and found that SRP-SF total scores were negatively associated with the reciprocity across all trials (*r*=-.27, *CI*[-.46, -.07]), across the ×4-trials (*r*=-.28, *CI*[-.46, -.07]), across the ×6-trials (*r*=-.30, *CI*[-.479, -.10]), but not the ×2-trials (*r*=-.13, *CI*[-.33 .08]). A complete overview of the results for the facet scores are displayed in Table 6.3.

Results of the linear mixed effects regression analysis are presented in Table 6.4. A main effect of investment was observed, such that higher investments lead to higher amounts of coins returned (β=.55, *t*(7650)=8.97, *p*<.001). Similarly, main effects were observed for multipliers ×2 (β=-15.61, *t*(7650)=-17.87, *p*<.001) and ×6 (β=-2.42, *t*(7650)=-2.75, *p*=.006) (both dummy-coded with reference condition x4). A significant interaction effect between investment and the multiplier of ×2 indicated that the differences in the amount of coins returned between multipliers ×2 and ×4 were larger for the higher investments (β=2.51, *t*(7650)=17.21, p<.001). Of the total sample of 86 participants, 83 Trustees returned non-zero amounts to the Investors. Thus, the HMTG successfully elicited reciprocity behavior.

6.3.2 Moral strategy model

The model parameters of the participants (phi, zeta) were distributed throughout the model's two-dimensional parameter space. This showed that different participants used different strategies to decide the amount of money to return to the Investor and thus confirms the importance of studying inter-participant variation in moral decision-making. The MSM was fitted to the behavioral response set of each participant in order to formally characterize the different strategies of the participants. All of the expected strategies were present in the sample (Figure 6.2, 5th column, and an overview of all participants is shown in Figure S6.1). The MSM classified 38 participants as inequity-averse, 36 participants as moral opportunists, 6 as greedy, 4 as generous, and 2 participants as guilt-averse (Figure 6.3).

6.3.2.1 Model performance

The performance of the MSM was compared with the component models of greed, guilt aversion and inequity aversion. The MSM performed significantly better than the other models, as determined by the Bayesian Information Criterion (BIC): ΔBIC with respect to greed model = –268.26, *t*(82) = –32.55, *p*<.001; ΔBIC w.r.t. guilt aversion = –79.12, *t*(82) = –12.28, *p*<.001; ΔBIC w.r.t. inequity aversion = –35.01, *t*(82) = –9.23, *p*<.001 (Figure 6.4). Parameter recovery tests indicated that the model was well identifiable, as the true parameters were recovered with very high accuracy (Pearson correlation between true and recovered parameter, phi: *r*(86) = 0.999, *p* < 0.001; zeta: *r*(86) = 0.998, *p* < 0.001) (Figure 6.5).

Table 6.3 Bayesian pairwise correlations between psychopathic traits and the reciprocity measures.

	×2 trials		×4 trials		×6 trials		all trials	
	Mode r	95% CI	Mode r	95% CI	Mode r	95% CI	Mode r	95% CI
SRP Interpersonal	-.10	[-.30, .11]	-.23*	[-.42, -.03]	-.24*	[-.42, -.03]	-.23*	[-.41, -.02]
SRP Affective	-.09	[-.29, .12]	-.22*	[-.41, -.01]	-.30*	[-.48, -.10]	-.24*	[-.43, -.04]
SRP Lifestyle	-.07	[-.27, .18]	-.15	[-.35, .06]	-.14	[-.34, .07]	-0.14	[-.34, .07]
SRP Antisocial	-.25*	[-.44, -.05]	-.40*	[-.56, -.21]	-.41*	[-.57, -.22]	-.40*	[-.56, -.21]
SRP Total	-.13	[-.33, .08]	-.28*	[-.46, -.07]	-.30*	[-.48, -.10]	-.27*	[-.46, -.07]

Notes. Mode r represents the most likely estimate for the strength of the correlation. Significant correlations are flagged with an asterisk (*).

6

Table 6.4 Results from the linear mixed effects regression for reciprocity behavior of the Trustee.

Fixed effects	Coefficients	SE	t	p
Intercept	13,68	0,61	22,419	<.001
Investment	0,55	0,061	8,972	<.001
Multiplier x2	-15,61	0,874	-17,871	<.001
Multiplier x6	-2,42	0,879	-2,752	0,006
Investment * Multiplier x2	2,51	0,146	17,212	<.001
Investment * Multiplier x6	0,21	0,143	1,48	-0,139
Random effects	**Term**		**SD**	
Subject	Intercept		4,493	

Notes. SE = standard error, SD=standard deviation

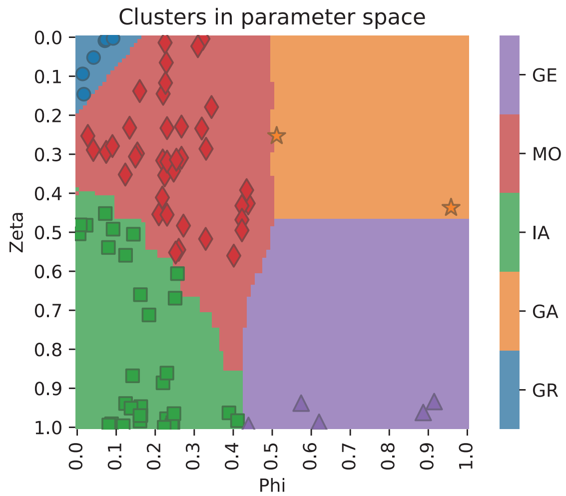


Figure 6.3 Distribution of participants in the phi-zeta model parameter space. Clusters are colour-coded. GE = generosity, MO = moral opportunism, IA = inequity aversion, GA = guilt aversion, and GR = greed.

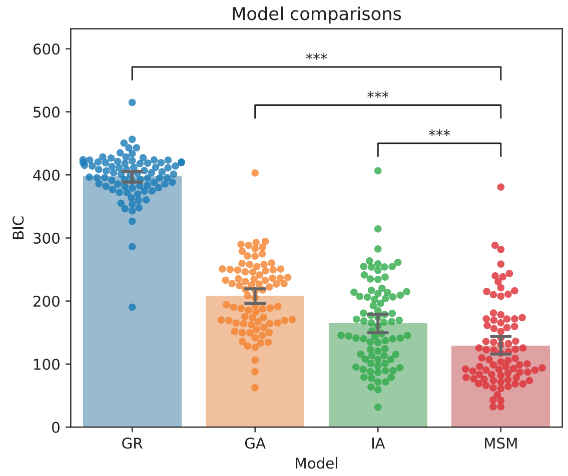


Figure 6.4 Model comparisons of the MSM with the unitary models of greed, guilt aversion and inequity aversion models. The BIC values are 397.41, 208.27, 164.16, and 129.15, respectively. A lower BIC value indicates a better model fit.

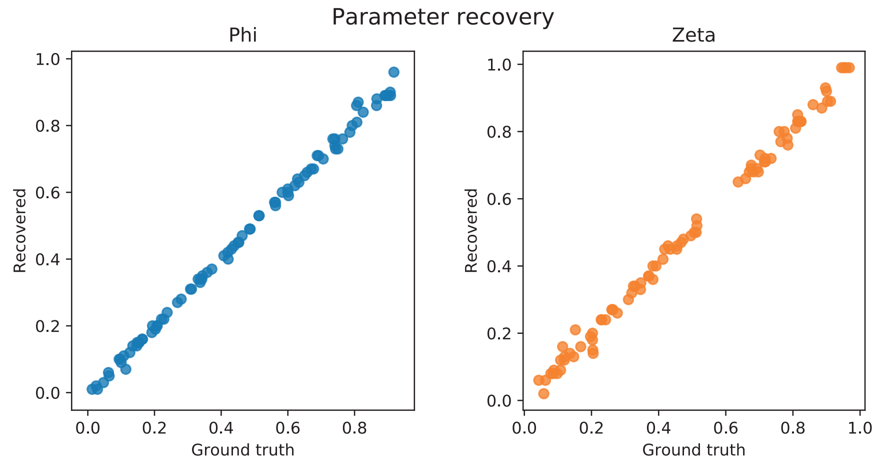


Figure 6.5 Parameter recovery tests show that the correspondence between the true and recovered parameters is high (phi: $r(86) = 0.999$, $p < 0.001$; zeta: $r(86) = 0.998$, $p < 0.001$).

Table 6.5 Bayesian pairwise correlations between SRP total and facet scores and the behavioral parameters.

	IA (zeta)		GA (phi)		Greed (theta)	
	Mode r	95% CI	Mode r	95% CI	Mode r	95% CI
SRP Interpersonal	-.16	[-.36, .04]	-.01	[-.22, .20]	-.14	[-.06, .34]
SRP Affective	-.28	[-.46, -.07]	-.04	[-.17, .25]	-.22	[.01, .41]
SRP Lifestyle	-.05	[-.26, .15]	-.07	[-.27, .14]	-.05	[-.14, .27]
SRP Antisocial	-.25	[-.43, -.04]	-.13	[-.33, .07]	-.25	[.06, .45]
SRP Total	-.21	[-.40, .00]	-.03	[-.24, .17]	.19	[-.02, .38]

Notes. Mode r represents the most likely mean, IA = inequity aversion, GA = guilt aversion.

6.3.2.2 Psychopathic traits

To investigate the relation between the moral strategies and psychopathic traits, we studied the correlation between the parameters phi and zeta on the one hand, and the SRP-SF total and facet scores on the other hand (Table 6.5). Although we did not find a significant correlation with the total score, there was a negative correlation between zeta and affective psychopathy traits (*r*=-.28, *CI*[-.46, -.07]) and antisocial traits (*r*=-.25, *CI*[-.43, -.04]). In contrast with our hypothesis, phi was not associated with the total score or any of the SRP subscales. Phi and zeta were not significantly correlated (*r*=.16, *CI*[-.05, .36]). Although Bayesian analysis is robust to outliers and heavy-tailed distributions (Gagnon, Desgagne, & Bedard, 2016), we repeated the analysis without the participant with the highest SRP score (*N*=85) as an additional check. The results showed that the effects remained unchanged (SRP affective: *r*=-.24, *CI*[-.42, -.03]; SRP antisocial: *r*=-.22, *CI*[-.41, -.01], see Table S6.1 for a complete overview of results). In addition, alternative tests using non-Bayesian Pearson and Spearman correlations yielded similar results (see Table S6.2).

Although phi and zeta were both negatively correlated with psychopathic traits, there was no significant correlation between the two. As an additional check, we performed a post hoc Hotelling's t-test with the cocor package in R (R Core team, 2013) to investigate whether the two correlations are significantly different from each other (Diedenhof & Musch, 2015). The results supported a difference between phi and zeta correlations with affective traits (*t*(83)=4.47, *p*=.016), but not antisocial traits (*t*(83)=1.17, *p*=.246).

6.4 Discussion

The aim of the present study was to elucidate moral strategies employed by a group of non-offenders with varying levels of psychopathic traits. Our modified version of the MSM is the first to additionally include generosity as a possible moral strategy. The distribution of the other moral strategies in the current sample is comparable with previous findings from a study that employed the MSM in healthy adults (van Baar et al., 2019; van Baar, Klaassen, Ricci, Chang, & Sanfey, 2020).

Our results indicate a negative relationship between psychopathic traits and reciprocity, as reflected by the amount of coins returned to the investor in the HMTG. On the facet level, it was shown that this association was significant for the interpersonal, affective, and antisocial traits. We suggest that findings could be linked to greed indirectly, where lower reciprocity is suggested to reflect greediness. The association between greed and psychopathic traits was also

6

proposed in previous literature. For instance, recent findings showed that antisocial individuals make choices based on the belief that most other individuals are as antisocial as they are themselves (Engelmann, Schmid, De Dreu, Chumbley, & Fehr, 2019), and use this belief to justify own behavior. This is in line with previous findings indicating that greedy individuals tend to engage more in making social comparisons (Krekels & Pandelaere, 2015). It was suggested that greed is partially rooted in the affective component of psychopathy (Rilling et al., 2007). The affective facet of psychopathy measures behaviors reflecting lack of empathy and failure to accept responsibility, among others. From this perspective, our results suggest that individuals scoring relatively high on the affective scale may care less about the negative consequences that their choices bring to others (Cima, Tonnaer, & Hauser, 2010), distance themselves from the consequences, and do not feel (entirely) responsible. Finally, previous findings showed that interpersonal traits, such as agreeableness, are propagated to maintain interpersonal harmony and promote prosocial behavior in socio-economic decision-making (Zhao & Smillie, 2015). A callous-conning interpersonal style on the other hand, would promote a more antisocial decision-making style. In the future, the expected increase in greedy behavior in individuals with psychopathic traits could be further investigated using novel computational approaches that would offer insight into the underlying strategies.

Our results with respect to zeta indicate that participants with elevated levels of affective or antisocial psychopathic traits were less averse to inequity in their social decisions, as was proposed in previous studies (Aharoni et al., 2011; Glenn, Raine, et al., 2009; Osumi & Ohira, 2010). One explanation could be that lower inequity aversion reflects disrupted affective responsiveness typically seen in individuals with elevated psychopathic traits, as affective responses are assumed to be crucial for judging the fairness of an offer (Barsky & Kaplan, 2007). If emotional processing is diminished, the ability to properly judge fairness may also be compromised as a consequence. This has implications for the development of morality, as impairments in affective processing are believed to bias moral development and result in antisocial behavior (Blair, 2013; Blair, 2007). The negative correlations between inequity aversion and the affective and antisocial components of psychopathy support this notion.

An alternative mechanism that has been suggested is that guilt aversion is the main driver of prosocial behavior. Although only two participants in our sample were categorized as purely guilt-averse, a large number of participants followed either a morally opportunistic (*N*=36) or generous strategy (*N*=4) strategy, in which guilt-averse behavior also plays a central role. In contrast with previous studies that emphasized the importance of guiltlessness in psychopathy (Koenigs, Kruepke, & Newman, 2010; Osumi & Ohira, 2017), the current study did not find

support for the link between psychopathic traits and GA. Recently, Gong et al. (2019) demonstrated that healthy individuals with high levels of psychopathic traits were able to understand the expectations of others, but did not seem to incorporate this information when making social decisions and experienced less aversion to anticipated guilt when deciding. The computational model for GA used by Gong et al. (2019) included a parameter that quantified how sensitive participants were to the experience of guilt as a consequence of letting their partner down (Chang et al., 2011; Charness & Dufwenberg, 2006; Dufwenberg & Gneezy, 2000; Khalmetski, 2016). Importantly, however, this GA model was not designed to untangle guilt aversion from inequity aversion, which is problematic as expectations often align with fairness norms (Chang et al., 2011; Sanfey, Stallem, & Chang, 2014; van Baar et al., 2019). This means that part of the results about guiltlessness and psychopathy in the literature could have been driven by the fact that inequity aversion and guilt aversion are confounded in many tasks (but not the HMTG). Therefore, we suggest that the previously proposed link between guilt aversion and psychopathy could be (partly) explained by decreased inequity aversion, which was not directly measured in prior studies (e.g. Gong et al., 2019; Seara-Cardoso et al., 2016).

Although phi and zeta were both negatively correlated with psychopathic traits (note that only the correlation with zeta was significant), there was no significant correlation between the two. As an additional check, we performed a post hoc test and found a significant difference between phi and zeta correlations with affective traits, but not antisocial traits. Based on these findings, we remain cautious with our interpretation concerning the association of antisocial traits and the computational parameters of inequity and guilt aversion. Nevertheless, our findings stress the importance of treating guilt and fairness as independent concepts, and it is possible that the inconsistencies pertaining the conceptualization of fairness and guilt could help explain the mixed results in the literature on psychopathy and moral decision-making. This could have important consequences for our view on psychopathy, which now assumes that guiltlessness is a main characteristic of individuals with high levels of psychopathic traits (Cima et al., 2010; Engelmann et al., 2019; Rilling et al., 2007; Krekels & Pandelaere, 2015).

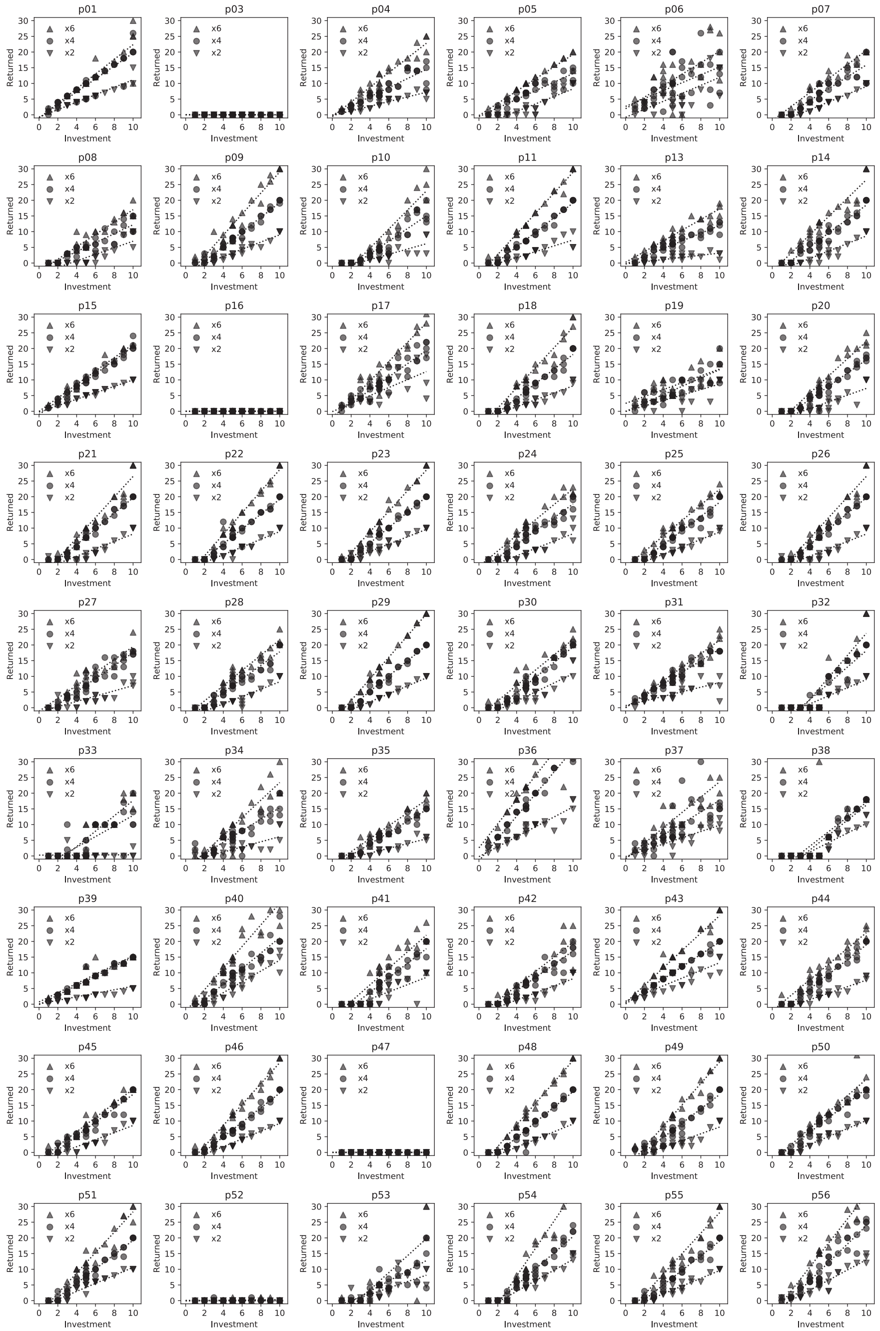
Furthermore, we found a substantial number of participants that switch between an IA and GA strategy, dependent on the context (MO: *N*=36, Generosity: *N*=4). Moral opportunists do not follow a 'pure' strategy but decide flexibly which course of action is both morally justifiable and maximally beneficial in a particular situation. Participants following a generous strategy ensure that guilt and inequity are always minimized. Given that MO and generosity were each defined as a combination of two parameters, there was no single model parameter that directly quantified these strategies. Therefore, we were unable to study the link between psychopathic traits, moral opportunism and generosity using the MSM model. Hopefully, further developments in the computational neuroscience will provide novel mathematical solutions that would allow us to quantify moral opportunism and generosity using single parameters individually capturing each strategy. Another solution could be to phenotype individuals based on their moral strategies in larger samples, and then examine the distribution of psychopathy scores across phenotypes. This could also have provided information about which clusters of personality and demographic traits were associated with each moral strategy (e.g. Driessen et al., 2018). This approach would be consistent with recent endeavors in computational psychiatry, in which computational parameters are used to assign individuals to homogeneous clusters, and then examining scores on behavioral and psychological variables (Brazil et al., 2017; Stephan, Iglesias, Heinzle, & Diaconescu, 2015). One of the strengths of our study was that participants were allowed to freely adopt their own moral strategy, in contrast to prior research where participants were instructed to reason in a particular way (e.g. Hein et al., 2016). However, the downside of this approach is that we were unable to control the prevalence of each moral strategy in our sample. This resulted in relatively low numbers of guilt-averse, greedy and generous participants. Therefore, we were unable to perform between-group analyses using the stratification of participants based on their moral strategy.

Another critical note is that the present study did not include a broad collection of additional measures to directly test the external and construct validity of the HMTG and the MSM, as this was not the goal of the present study. Instead, we built upon other work in which extensive validation is offered for the moral strategies identified using the MSM by referencing the strategies to an array of external task and questionnaire measures. Importantly, it was demonstrated that the moral strategies align with participant's own experience of their psychological decision process (van Baar et al., 2020). These results supported the external and construct validity of the task and the model.

In conclusion, the current study unraveled five distinct cognitive strategies underlying social decision-making, and examined the relationships of guilt, inequity, and greed with psychopathic traits. Our results indicated that the affective traits that are associated with psychopathy were linked to lower levels of reciprocity, which indirectly reflects greed. Our computational analyses provided more insight into the strategies that underlie reduced reciprocity behavior in individuals with a high level of affective traits. They indicated that a reduced sense of fairness, but not guiltlessness underlie the reduced reciprocity behavior in individuals with high levels of affective traits. This could explain the link between affective traits and poor social decision-making and immoral behavior in individuals with elevated psychopathic traits. In contrast with earlier results, our findings could

not confirm lack of guilt as a key factor in social decision-making in individuals with psychopathic tendencies. Our findings stress the importance of treating guilt and fairness as independent concepts, and it is possible that the lack of methodological and conceptual precision in untangling the individual impact of fairness and guilt in previous studies could explain the mixed results in moral decision-making literature. Obtaining a fine-grained perspective on moral strategies could open new avenues for research in offender populations as well, which could eventually boost the development of treatment that offers a better fit with the cognitive capacities of such individuals (Baskin-Sommers et al., 2015).

6.5 Supplementary material



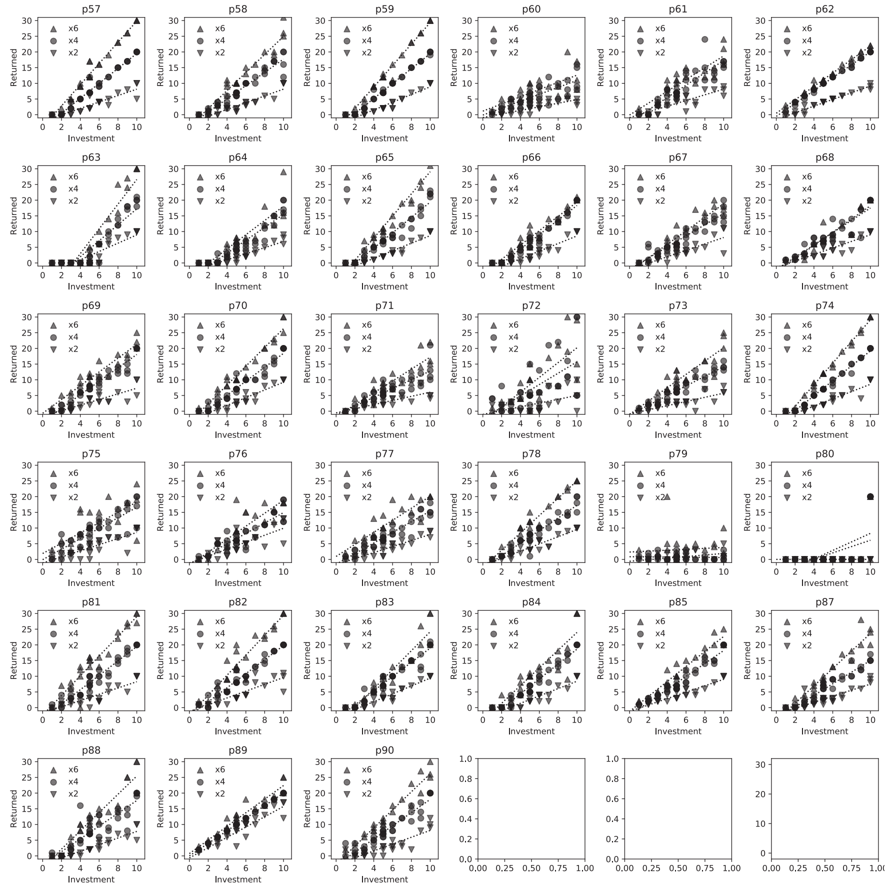


Figure S6.1 Choice behaviour of each participant. Grey symbols reflect the returned amount of coins on each trial (black symbols result from an overlay of two or multiple grey symbols).

Table S6.1 Bayesian pairwise correlations between computational parameter and SRP scores without highest score SRP score (n=85).

	Phi (guilt aversion)		Zeta (inequity aversion)	
	Mode r	95% CI	Mode r	95% CI
SRP Interpersonal	.01	[-.20, .21]	-.15	[-.35, .06]
SRP Affective	.10	[-.11, .30]	-.24*	[-.42, -.03]
SRP Lifestyle	-.04	[-.25, .17]	-.03	[-.23, .18]
SRP Antisocial	-.05	[-.26, .16]	-.22*	[-.41, -.01]
SRP Total	-.01	[-.20, .22]	-.16	[-.36, .05]

Notes. Mode r = most likely estimate of the strength of the correlation, Significant correlations are flagged with an asterisk (*).

Table S6.2 Pearson and Spearman correlations between SRP scores and computational parameters

	Phi (guilt aversion)		Zeta (inequity aversion)	
	r	rho	r	rho
SRP Interpersonal	-.01	-.03	-.17	-.18
SRP Affective	.04	.05	-.28**	-.25*
SRP Lifestyle	-.07	.04	-.05	-.07
SRP Antisocial	-.14	-.06	-.25*	-.25*
SRP Total	-.04	<-.01	-.21	-.19

Notes. Significant correlations are flagged with an asterisk (*). Rho = Spearman's correlation coefficient, r = Pearson's correlation coefficient.

7

General discussion

The main aim of this thesis was to further unravel the nature of the mechanisms that are believed to play a role in disruptions in affective processing and decision-making seen in relation with high levels of psychopathic traits. Behavioral, electro-physiological and computational approaches were applied to study affective and social cognitive processes in a series of experiments. Firstly, I will provide a summary of the results included in this thesis. Secondly, I will integrate the main findings and discuss the results in the light of the IES model. Finally, I will discuss potential future directions that would be beneficial to the field of psychopathy and associated mechanisms of decision-making, followed by some concluding remarks.

7.1 Summary

Antisocial behavior is a heterogeneous construct and there is ample evidence that supports the existence of different subtypes of antisocial individuals (e.g. see Brazil et al., 2018). Subtyping of antisocial individuals is often theory-driven and based on time-consuming measures, such as the Psychopathy Checklist Revised (PCL-R). Chapter 2 described a study in which latent profile analysis was performed using the Self-Report Psychopathy Checklist Short Form (SRP-SF) to identify anti-social profiles in male offenders. In addition, we studied how these profiles were linked to personality correlates and a broad range of behaviours seen in antisocial populations. The experiment yielded extensive and multifaceted characterizations of four different profiles; generic offenders, impulsive-antisocial traits offenders, non-antisocial psychopathic traits offenders, and psychopathic traits offenders. These results were in line with previous subtyping studies that were based on the PCL-R. Taken together, these results provided support for the presence of different antisocial personality profiles in an offender population and for the validity of the SRP-SF as a measure to subtype individuals based on their level of psychopathic traits.

In the following two chapters, social-affective functioning was examined in relation to psychopathic traits. The objective was to obtain a deeper understanding of the link between psychopathic traits and the processing of threat cues in a community sample. In chapter 3, a study is described in which we examined the automatic approach and avoidance responses to emotional facial expressions in a sample of cognitively unimpaired adults. In most people, confrontation with a threatening stimulus, e.g. an angry face, elicits personal distress and initiates an avoidance response (Lang et al., 1997). As a recent study suggested reduced threat avoidance in psychopathic offenders, we were interested in the automatic response towards angry facial expressions specifically, and whether this effect was associated with the level of psychopathic traits in a community sample. More importantly, we aimed to further unravel this threat coping mechanism by examining the role of testosterone, given its link with aggression (Book et al., 2001; Popma et al., 2007) and threat approach (Hermans, Ramsey, & van Honk, 2008; Radke et al., 2015; Terburg & van Honk, 2013). Participants were included based on their SRP total score, in order to obtain a reliable distribution of psychopathic traits in our test sample. We found that the level of psychopathic traits was linked to reduced automatic threat avoidance. The results suggested that individuals with high levels of psychopathic traits do not show the typical avoidance reaction towards angry facial expressions, but instead show an approach reaction. This finding was contistent with a previous study demonstrating a similar effect in psychopathic offenders (von Borries et al., 2012). Our results indicated that although testosterone was positively related to psychopathic traits, it did not mediate the effect of psychopathic traits on threat avoidance. However, as was proposed in previous studies (Dabbs et al., 1991; Popma et al., 2007), it could be that cortisol could has moderated the behavioral effects of testosterone. Therefore, it is important to consider the interplay between the Hypothalamic-Pituitary-Adrenal (HPA) and testosterone in future studies.

The study described in chapter 4 investigated pain empathy and pain sensitivity in relation to psychopathy by measuring event-related potentials (ERPs) extracted from the ongoing EEG in an interactive setup. Each participant first fulfilled the role of "villain" (observing another person receiving electronic shocks) and later of "victim" (receiving the shocks while another person is watching). In addition, control over the painful stimulus was modulated, where "passive" refers to having no control over the shocks, while "active" refers to having control over delivering the shocks. This resulted in four different conditions; passive villain, active villain, active victim, and passive victim. Response-, visual- and pain ERPs were compared between the four conditions. The findings suggested that individuals experienced more conflict when hurting someone else than when hurting themselves. In contrast with the hypothesis, this effect was found to be independent of the level of psychopathic traits. Furthermore, the results indicated that self-controlled pain was experienced as more painful than uncontrolled pain, and this effect was found to be negatively related to the level of psychopathic traits. It was proposed that this relationship could be explained by the reduced sensitivity to pain that was found in previous studies (Brislin, Buchman-Schmitt, Joiner, & Patrick, 2016; Hare, 1965). In sum, the results of this study suggested that social context, attention and personality traits are important modulators of pain- and empathy-related neuronal responses. However, taking the studies limitations into account, additional research is required to validate the effects and take a closer look at the individual moderators of pain processing.

Chapters 5 and 6 describe two studies that examined how different mechanisms involved in decision-making are related to the level of psychopathic traits. The study in chapter 5 aimed to investigate how individuals make use of social and non-social information in a reinforcement learning task, in which the trade-off between two types of information affects task performance and associated monetary reward for the participant. More specifically, the latent cognitive processes that are involved in associative learning of stable and volatile information were studied and the effect of psychopathic traits was investigated. In addition, oscillatory theta activity was examined, given its potential involvement in adaptive control processes. The findings indicated that individuals with psychopathic traits experienced impairments in associative learning based on social information, and suggested that interpersonal traits are linked to a reduced ability to adapt to changes in the reliability of social information. In addition, it was found that decreased theta power was linked to higher levels of psychopathic traits, which aligns with indications that theta is involved in tracking the volatility of social information (Behrens et al., 2007). Furthermore, the impairments that were associated to increased levels of psychopathic traits, mostly present in the social domain, did not lead to a preference for one of the information sources, and did not affect the risk that was taken in order to obtain a high reward.

Chapter 6 describes a study in which moral strategies (i.e. decision styles) were investigated by assessing reciprocity in a socio-economic trust game. In this task, choice behavior of the participant affected both the participant and the confederate. A computational model was used to estimate the role of different strategies in this task and examined how this was related to psychopathic traits. These results indicated that there were five different moral strategies used in this task; guilt-aversion, inequity aversion, moral opportunism, greed, and generosity. Two parameters representing inequity aversion and guilt aversion were measured in relation to the level of psychopathic traits of the participants. Inequity aversion in this task reflected the tendency to ensure an even split between the participant and the confederate. Guilt aversion reflected the tendency to follow other's expectations on order to avoid guilt feelings. The results showed that inequity aversion, but not guilt aversion, was associated with lower levels of psychopathic traits. Furthermore, our non-modelling results indicated that higher psychopathic traits were associated with lower reciprocity, that is the amount of coins returned to the confederate. This could suggest that individuals with high levels of psychopathic traits act greedier in a socio-economic game, which is consistent with previous findings in literature. Based on these findings, it was suggested that the underlying mechanism of greedy behavior in these individuals could be reflected by a decreased sense of fairness.

7.2 Integration of the key findings

The first empirical study described in this thesis (chapter 2) presented four different antisocial subtypes in an offender population. These profiles that were based on self-report measures further emphasized the importance of treating psychopathy as a multidimensional construct. Before continuing with a discussion on the main findings of the other studies that were presented in this thesis, I would like to clarify why we did not look into subtypes in the general population. Psychopathic traits are normally distributed among the population and criminal psychopaths are believed to express these traits to an extreme extent. This means that the different antisocial subtypes in the community are typically less pronounced. Therefore, a large sample is necessary to expose these subtypes. Unfortunately, we did not have the resources to recruit a sufficiently large sample to be able to perform a latent profile analysis and obtain meaningful subtypes. Instead we applied an oversampling procedure to select participants for three of the studies that were described in this thesis. More specifically, the lower and higher ends of the distribution of psychopathic traits were oversampled in order to enhance the presence of extreme scores on both sides of the distribution. Importantly, this method has been proven to be successful in previous studies (Bernat et al., 2011; Brazil, Maes et al., 2013; Gong et al., 2019). Furthermore, the results of our subtyping study emphasized the importance of treating psychopathy as a multifaceted construct. An extensive body of literature has confirmed that psychopathy is made up of at least four dimensions reflecting interpersonal, affective, lifestyle and antisocial deviations. These four facets are fundamentally interrelated (Hare & Neumann, 2005; Vitacco et al., 2005) and evidence across different samples showed that a superordinate factor could account for most of the variance of the four dimensions (Neumann et al., 2007). Therefore, in all of our studies, we investigated the four facets individually, as well as the superordinate factor that was represented by the total score of the SRP-SF.

In general, findings that were described in this thesis are consistent with the common thread in psychopathy research indicating emotional impairments that manifest itself in poor decision-making patterns. Chapters 3 and 4 specifically examined social-affective functioning in relation to psychopathic traits. The two studies differed in many ways, including the affective modality that was measured, the social context that participants were exposed to, the study environment, and the type of measurements that were applied. However, the results of both studies suggested that individuals with high levels of psychopathic traits show abnormal processing of affective stimuli that are associated with threat. Confrontation with an angry facial expression typically results in an avoidance response, presumably because a person wants to avoid threat to him- or herself. This effect was absent in both offenders (von Borries et al., 2012) and community-dwelling individuals with high levels of psychopathic traits (chapter 3), and instead, the results suggested that individuals with high levels of psychopathic traits tended to approach a threatening stimulus. Besides, studies investigating the link between psychopathy and the processing of pain suggested that individuals characterized by high levels of psychopathic traits showed reduced fear in response to anticipated pain (Hare, 1965), and showed an increased pain tolerance (Brislin et al., 2016). The findings in chapter 4 also indicated an abnormal response towards imminent pain, therewith supporting the hypothesis of reduced threat avoidance in relation to psychopathic traits. Finally, although there was some variability between the two studies, the effect was significant on both facet and superordinate level.

The IES model suggests that psychopathic individuals experience impaired processing of distress cues (e.g. fear, pain, sadness) in others (Blair, 2013). However, the model does not distinguish emotions in others that could potentially be harmful to oneself (e.g. anger, threat) from emotions that do not directly affect oneself but typically underlie empathic responding (e.g. sadness, pain). While the results of this thesis supported the evidence indicating an association between psychopathy and abnormal threat reactivity, the neurophysiological correlate reflecting conflict in response to pain experienced by others was not different in individuals scoring high on psychopathic traits. Concerning the latter, it should be noted that we examined the neurological response that is associated with conflict and used it to estimate empathy. It would have been informative to explicitly ask them how they felt when watching someone else receiving the shock.

The two other empirical studies that were described in this thesis investigated mechanisms of decision-making and examined the effect of psychopathic traits on these mechanisms. The results of these studies emphasized the impact of psychopathic traits on choice behavior in social situations. The findings of chapter 5 suggested that while learning from reward-based information was intact, learning from social information was negatively associated with psychopathic traits. Interpersonal traits were linked to a reduced ability to adapt to changes in the reliability of social information. Although this affected their performance in the task, it did not change their level of risk-taking. The findings of this study highlight the relationship between impaired social-affective functioning and its effect on decision-making in a social context. Interestingly, the impaired processing of social and affective information does not necessarily result in restrained behavior or a passive attitude. In fact, regardless of its effect on performance/success, it seems that these individuals do not avoid risky decisions that could have a negative impact on themselves or others. This was also reflected in the results of chapter 6, which showed that psychopathic traits were positively associated with greed. This finding suggested that an individual with a tendency to psychopathy keeps most, if not all, of the reward for him- or herself, even if this is at the expense of another person and bears a risk of losing face, which could be disadvantageous in future negotiations. This interpretation could be linked to Gray's reinforcement sensitivity theory (Gray, 1970), which was already introduced in the first chapter of this thesis. Based on earlier findings, it was suggested that interpersonal-affective traits were linked to an underactive behavioral inhibition system (BIS), while impulsive-antisocial traits were linked to an overactive behavioral activation system (BAS; Fowles, 1980; Newman et al., 2005; Wallace et al., 2009). The approach tendency towards threatful stimuli that was discussed in chapter 3 could also be a result of the disbalance of the two systems. Following up on this theory, Gray proposed that a strong BAS might be associated with a relatively automatic bias to attend to goal-relevant cues at the expense of processing cues that are peripheral to one's dominant response set. This reflects the idea that individuals with high levels of lifestyle-antisocial psychopathic traits are characterized by a tendency to excessively pursue appetitive stimuli and is consistent with evidence indicating that individuals scoring high on impulsive and antisocial traits are hypersensitive to reward (Buckholtz et al., 2010; Fowles, 1988; Moltó, Poy, Segarra, Pastor, & Montañés, 2007; Stoff, Breiling, & Maser, 1997). This tendency has been associated with a decreased ability to properly monitor and adjust ongoing behavior in response to changes in environmental contingencies (Buckholtz et al., 2010; Fowles, 1980; Newman, 1987; Newman et al., 2005; Newman et al., 1997; Hoppenbrouwers, Neumann, Lewis, & Johansson, 2015). Therefore, this hypothesis lends itself also to explain the perserverance in using social information and the unchanged level of risk-taking in individuals with high levels of psychopathic traits, even though performance was reduced due to a diminished ability to adapt to changes in the reliability of social information.

According to the IES model, the core decision-making deficit in individuals with high levels of psychopathic traits are suggested to be associated impairments in the vmPFC and the striatum. The link with the vmPFC was supported by our findings described in chapter 5 indicating the association between impairments in adaptive control during associative learning of social information and reduced midfrontal theta oscillations. The overactive behavioral activation system and the heightened responsivity to reward that was proposed to underlie some of our findings in three of the chapters in my thesis (3, 5 and 6) may explain the link between psychopathy and the striatum. Buckholtz and colleagues (2010) found that some of the behavioral correlates that are thought to play a role in the heightened sensitivity to reward, such as impulsive and antisocial behavior, are caused by neurochemical and neurophysiological overreactivity of the dopaminergic reward system, including the ventral striatum.

7.3 Future directions

In the next paragraph, I would like to highlight two topics that were addressed in this thesis and discuss their potential in future research to further develop our understanding of psychopathy, but also psychiatric disorders in a broader sense.

Although psychopathy is often viewed as a unitary construct, a large body of scientific evidence emphasized its complex heterogeneous construct. Furthermore, it was shown that the personality features that are associated with the disorder vary along continuous dimensions. The combination of the presence of specific traits determine the subtype of a particular person. Since the personality traits are typically more pronounced in clinical samples, it is also easier to detect subtypes in these samples. To date, there are several studies that investigated antisocial subtypes based on psychopathy measures (Hare, 2016; Krstic et al., 2018; Mokros et al., 2015; Skeem, Johansson, Andershed, Kerr, & Louden, 2007), including the study that was described in chapter 2 (Driessen et al., 2018). These studies provided important insight into the different antisocial profiles and their personality and behavioral characteristics. However, the current lack of successful treatment possibilities for individuals with antisocial and psychopathic traits demonstrates the limited understanding we have about the underlying causes of the behavior and how this relates to the different subtypes. Over the years, there have been great developments in the field of neuroscience that have contributed to our awareness of the impaired cognitive processes and involved brain regions. Similarly, the field of biology has brought insight into the biological characteristics of the disorder. Nevertheless, the number of different neurocognitive and neurobiological accounts on psychopathy demonstrates the complexity and heterogeneity of the construct. By combining insights and methodological approaches from diverse research fields and by applying subtyping not only on behavioral characteristcs, but also on these other measures, we could provide a more complete overview of the multidisciplinairy characteristics of different antisocial and psychopathic subtypes (Brazil et al., 2018). Eventually, this information could be used to develop personalized and targeted treatment programs. Although the study of subtypes in the general population is more challenging, it could provide interesting insights into the differences and similarities between individuals with psychopathic traits that are successful and unsuccessful in society.

As was addressed earlier in this thesis, the need for further specification of the cognitive processes that are impaired in psychopathy is necessary to improve subtyping of antisocial individuals. While traditional behavioral and neuroimaging studies have increased our understanding of the *what* (e.g. what is psychopathy) and *where* (e.g. where in the brain, i.e. what brain regions are involved) questions regarding psychopathy, computational approaches may help us to find an answer to the *how* (e.g. how does it work, i.e. what mechanisms are involved/impaired) questions. The use of computational models to study behavioral mechanisms has gained popularity within psychiatry research in general, and has already led to many new insights into the neurobehavioral mechanisms that underlie several psychiatric disorders (Maia & Frank, 2011; Mars et al., 2012; Paulus, Huys, & Maia, 2016; Schmidt et al., 2011; Stephan & Mathys, 2014). However, the implementation of computational modelling approaches in the field of psychopathy research is limited. Only a few studies, including the ones described in chapter 5 and 6, used such models investigate reinforcement learning in relation to psychopathy (Blair, 2004; Brazil et al., 2017; Brazil, Maes, et al., 2013; Oba et al., 2019). This is unfortunate, as these models allow us to obtain a detailed view on the sources underlying the cognitive impairments that are associated with psychopathy, while traditional behavioral and neuroimaging findings do not provide insight into these latent cognitive operations. We could use this knowledge to stratify individuals based on these unified underlying mechanisms. Furthermore, we could utilise model- based approaches for the subtyping itself. That is, instead of categorizing subtypes based on theoretical accounts, we could use computational models to identify latent variables that characterize clusters in the data. Bridging of multiple levels of analysis could provide a more complete view on psychopathy and lead to better classification accuracy. The next step would be to validate the subtypes and to test whether the results obtained with these approaches improve the physician's ability to predict future clinical outcomes and select optimal treatments for individual patients.

Finally, I would like to highlight a recently emerging approach that was proposed to complement the current subtyping approaches. The modelling techniques that were discussed earlier allow for obtaining a deeper insight into the underlying mechanisms that are associated with the cognitive impairments of psychopathy. In addition, they can be used to improve our classification accuracy to obtain meaningful subtypes of antisociality. However, a general criticism on subtyping techniques is that clustering is focused on group averages and ignores individual variations within clusters. Normative modelling is a statistical approach that aims to delineate individual variations of psychiatric disorders across different dimensions. More specifically, it provides statistical inferences that represents the degree to which each individual deviates from the normative pattern. This way, normative modelling can be used to understand the variation across the population independently of the clinical labels (a detailed review on the use of normative models in psychiatry is provided by Marquand, Kia, Zahibi, Wolfers, Buitelaar, & Beckmann, 2019). For a dimensional and heterogeneous construct such as psychopathy, normative modelling on large population samples could shed light on variations in brain and behavior that potentially discriminate successful and unsuccessful adaptation to society.

7.4 Concluding remarks

The work presented in this thesis offers novel insights into potential mechanisms underlying social learning and decision-making and demonstrates how these mechanisms may be affected by psychopathic personality traits. Furthermore, it highlights the importance of improving research techniques that allow us to investigate the neurocognitive processes that underlie psychopathic-like behaviors in detail and to integrate findings and perspectives across different research areas.

References

A

Aharoni, E., Antonenko, O., & Kiehl, K. A. (2011). Disparities in the moral intuitions of criminal offenders: The role of psychopathy. *Journal of Research in Personality, 45*(3), 322-327.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.).

Anderson, N. E., & Kiehl, K. A. (2012). The psychopath magnetized: insights from brain imaging. *Trends in cognitive sciences, 16*(1), 52-60.

Archer, J., & Webb, I. A. (2006). The relation between scores on the Buss–Perry Aggression Questionnaire and aggressive acts, impulsiveness, competitiveness, dominance, and sexual jealousy. A*ggressive Behavior: Official Journal of the International Society for Research on Aggression, 32*(5), 464-473.

Arnett, P. A., Smith, S. S., & Newman, J. P. (1997). Approach and avoidance motivation in psychopathic criminal offenders during passive avoidance. *Journal of personality and social psychology, 72*(6), 1413-1428.

Arntz, A., & Hopmans, M. (1998). Underpredicted pain disrupts more than correctly predicted pain, but does not hurt more. *Behav Res Ther, 36*(12), 1121-1129.

Arrigo, B. A., & Shipley, S. (2001). The confusion over psychopathy (I): Historical considerations. *International Journal of Offender Therapy and Comparative Criminology, 45*(3), 325-344.

Avenanti, A., Bueti, D., Galati, G., & Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nat Neurosci, 8*(7), 955-960. doi:10.1038/nn1481

B

Babiak, P., Hare, R. D., & McLaren, T. (2006). *Snakes in suits: When psychopaths go to work*: Regan Books New York, NY.

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the Association between Latent Class Membership and External Variables Using Bias-adjusted Three-step Approaches. *Sociological Methodology, 43*(1), 272-311. doi:10.1177/0081175012470644

Barsky, A., & Kaplan, S. A. (2007). If you feel bad, it's unfair: A quantitative synthesis of affect and organizational justice perceptions. *Journal of applied psychology, 92*(1), 286.

Baskin-Sommers, A. R., Curtin, J. J., & Newman, J. P. (2015). Altering the Cognitive-Affective Dysfunctions of Psychopathic and Externalizing Offender Subtypes with Cognitive Remediation. *Clinical Psychological Science, 3*(1), 45-57. doi:10.1177/2167702614560744

Baskin-Sommers, A. R., Wallace, J. F., MacCoon, D. G., Curtin, J. J., & Newman, J. P. (2010). Clarifying the factors that undermine behavioral inhibition system functioning in psychopathy. *Personality Disorders: Theory, Research, and Treatment, 1*(4), 203.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and social psychology bulletin, 23*(7), 751-758.

Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American economic review, 97*(2), 170-176.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological methods, 9*(1), 3.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: an interpersonal approach. *Psychological bulletin, 115*(2), 243.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58*(4), 639-650.

Behrens, T. E., Hunt, L. T., & Rushworth, M. F. (2009). The computation of social behavior. *science, 324*(5931), 1160-1164.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature, 456*(7219), 245-249.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience, 10*(9), 1214-1221.

Bernat, E. M., Nelson, L. D., Steele, V. R., Gehring, W. J., & Patrick, C. J. (2011). Externalizing psychopathology and gain–loss feedback in a simulated gambling task: Dissociable components of brain response revealed by time-frequency analysis. *Journal of Abnormal Psychology, 120*(2), 352.

Birbaumer, N., Veit, R., Lotze, M., Erb, M., Hermann, C., Grodd, W., & Flor, H. (2005). Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. *Archives of general psychiatry, 62*(7), 799-805.

Blackburn, R. (2007). Personality disorder and antisocial deviance: comments on the debate on the structure of the psychopathy checklist-revised. *Journal of personality disorders, 21*(2), 142-159.

Blair, R. (1995). A cognitive developmental approach to morality: Investigating the psychopath, in «Cognition», 57.

Blair, R. (2003). Facial expressions, their communicatory functions and neuro–cognitive substrates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1431), 561-572.

Blair, R. (2004). The roles of orbital frontal cortex in the modulation of antisocial behavior. *Brain and cognition, 55*(1), 198-208.

Blair, R. (2005). Applying a cognitive neuroscience perspective to the disorder of psychopathy. *Development and psychopathology, 17*(3), 865-891.

Blair, R. (2013). The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience, 14*(11), 786-799.

Blair, R., Mitchell, D., Leonard, A., Budhani, S., Peschardt, K., & Newman, C. (2004). Passive avoidance learning in individuals with psychopathy: Modulation by reward but not by punishment. *Personality and Individual Differences, 37*(6), 1179-1192.

Blair, R. J. (2012). Cortical thinning and functional connectivity in psychopathy. *The American Journal of Psychiatry, 169*(7), 684-687. doi:10.1176/appi.ajp.2012.12030396

Blair, R. J., & Cipolotti, L. (2000). Impaired social response reversal: A case ofacquired sociopathy'. *Brain, 123*(6), 1122-1141.

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in cognitive sciences, 11*(9), 387-392.

Blair, R. J. R., Peschardt, K. S., Budhani, S., Mitchell, D., & Pine, D. (2006). The development of psychopathy. *Journal of child psychology and psychiatry, 47*(3-4), 262-276.

Boccardi, M., Frisoni, G. B., Hare, R. D., Cavedo, E., Najt, P., Pievani, M., . . . Repo-Tiihonen, E. (2011). Cortex and amygdala morphology in psychopathy. *Psychiatry Research: Neuroimaging, 193*(2), 85-92.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American economic review, 90*(1), 166-193.

Book, A. S., & Quinsey, V. L. (2004). Psychopaths: cheaters or warrior-hawks? *Personality and individual differences, 36*(1), 33-45. doi:https://doi.org/10.1016/S0191-8869(03)00049-7

Book, A. S., Starzyk, K. B., & Quinsey, V. L. (2001). The relationship between testosterone and aggression: A meta-analysis. *Aggression and Violent Behavior, 6*(6), 579-599.

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage, 25*(1), 312-319.

Brandt, J. R., Kennedy, W. A., Patrick, C. J., & Curtin, J. J. (1997). Assessment of psychopathy in a population of incarcerated adolescent offenders. *Psychological Assessment, 9*(4), 429.

Brazil, I., Hunt, L., Bulten, B., Kessels, R., De Bruijn, E., & Mars, R. (2013). Psychopathy-related traits and the use of reward and social information: a computational approach. *Frontiers in psychology, 4*, 952.

Brazil, I., Mathys, C., Popma, A., Hoppenbrouwers, S., & Cohn, M. (2017). Representational uncertainty in the brain during threat conditioning and the link with psychopathic traits. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2*(8), 689-695.

Brazil, I. A., & Cima, M. (2016). 10 Contemporary approaches to psychopathy. *The handbook of forensic psychopathology and treatment*, 206.

Brazil, I. A., Hunt, L. T., Bulten, B. H., Kessels, R. P., De Bruijn, E. R., & Mars, R. B. (2013). Psychopathy-related traits and the use of reward and social information: a computational approach. *Frontiers in psychology, 4*, 952.

Brazil, I. A., Maes, J. H., Scheper, I., Bulten, B. H., Kessels, R. P., Verkes, R. J., & de Bruijn, E. R. (2013). Reversal deficits in individuals with psychopathy in explicit but not implicit learning conditions. *Journal of psychiatry & neuroscience: JPN, 38*(4), E13.

Brazil, I. A., Mars, R. B., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., & De Bruijn, E. R. (2011). A neurophysiological dissociation between monitoring one's own and others' actions in psychopathy. *Biological psychiatry, 69*(7), 693-699.

Brazil, I. A., van Dongen, J. D., Maes, J. H., Mars, R., & Baskin-Sommers, A. R. (2018). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews, 91*, 259-277.

Brislin, S. J., Buchman-Schmitt, J. M., Joiner, T. E., & Patrick, C. J. (2016). "Do unto others"? Distinct psychopathy facets predict reduced perception and tolerance of pain. *Personality Disorders: Theory, Research, and Treatment, 7*(3), 240.

Bromm, B. (1995). Consciousness, pain, and cortical activity. *Pain and the brain: from nociception to cognition*, 35-39.

Brook, M., Brieman, C. L., & Kosson, D. S. (2013). Emotion processing in Psychopathy Checklist—Assessed psychopathy: A review of the literature. *Clinical psychology review, 33*(8), 979-995.

Brosnan, S. F. (2006). Nonhuman species' reactions to inequity and their implications for fairness. *Social Justice Research, 19*(2), 153-185.

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Benning, S. D., Li, R., . . . Shelby, E. S. (2010). Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. *Nature neuroscience, 13*(4), 419-421.

Bufalari, I., Aprile, T., Avenanti, A., Di Russo, F., & Aglioti, S. M. (2007). Empathy for pain and touch in the human somatosensory cortex. *Cerebral Cortex, 17*(11), 2553-2561.

Burns, S., Roberts, L. D., Egan, S., & Kane, R. (2015). Evaluating emotion processing and trait anxiety as predictors of non-criminal psychopathy. *Personality and individual differences, 81*(Supplement C), 148-154. doi:https://doi.org/10.1016/j.paid.2014.08.044

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in cognitive sciences, 4*(6), 215-222.

C

Carré, J. M., Fisher, P. M., Manuck, S. B., & Hariri, A. R. (2012). Interaction between trait anxiety and trait anger predict amygdala reactivity to angry facial expressions in men but not women. *Social Cognitive and Affective Neuroscience, 7*(2), 213-221. doi:10.1093/scan/nsq101

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of personality and social psychology, 67*(2), 319.

Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2012). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral cortex, 22*(11), 2575-2586.

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage, 49*(4), 3198-3209.

Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology, 49*(2), 220-238.

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron, 70*(3), 560-572.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica, 74*(6), 1579-1601.

Chen, A. C., Chapman, C. R., & Harkins, S. W. (1979). Brain evoked potentials are functional correlates of induced pain in man. *Pain, 6*(3), 365-374.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and social psychology bulletin, 25*(2), 215-224.

Christensen, R. (2019). Covariance Parameter Estimation. In *Advanced Linear Modeling* (pp. 125-159): Springer.
Christianson, S.-Å., Forth, A. E., Hare, R. D., Strachan, C., Lidberg, L., & Thorell, L.-H. (1996). Remembering details of emotional events: A comparison between psychopathic and nonpsychopathic offenders. *Personality and individual differences, 20*(4), 437-443.
Cima, M., Tonnaer, F., & Hauser, M. D. (2010). Psychopaths know right from wrong but don't care. *Social cognitive and affective neuroscience, 5*(1), 59-67.
Clark, A. P., Bontemps, A. P., Batky, B. D., Watts, E. K., & Salekin, R. T. (2019). Psychopathy and neurodynamic brain functioning: A review of EEG research. *Neuroscience & Biobehavioral Reviews, 103*, 352-373.
Cleckley, H. M. (1964). *The mask of sanity: An attempt to clarify some issues about the so-called psychopathic personality*: Ravenio Books.
Cohen, M. X. (2008). Neurocomputational mechanisms of reinforcement-guided learning in humans: a review. *Cognitive, Affective, & Behavioral Neuroscience, 8*(2), 113-125.
Cohen, M. X., Elger, C. E., & Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage, 35*(2), 968-978.
Colins, O. F., Fanti, K. A., Salekin, R. T., & Andershed, H. (2017). Psychopathic personality in the general population: Differences and similarities across gender. *Journal of Personality Disorders, 31*(1), 49-74.
Cooke, D. J., & Michie, C. (1999). Psychopathy across cultures: North America and Scotland compared. *Journal of abnormal psychology, 108*(1), 58-68.
Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment, 13*(2), 171.
Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in cognitive sciences, 10*(7), 294-300.
Crego, C., & Widiger, T. A. (2015). Psychopathy and the DSM. *Journal of personality, 83*(6), 665-677.

D

Dabbs, J. M., Jurkovic, G. J., & Frady, R. L. (1991). Salivary testosterone and cortisol among late adolescent male offenders. *Journal of abnormal child psychology, 19*(4), 469-478.
Davis, M., & Whalen, P. J. (2001). The amygdala: vigilance and emotion. *Molecular psychiatry, 6*(1), 13-34.
Davis, M. H. (1980). A multidimensional approach to individual differences in empathy.
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.
Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature neuroscience, 3*(11), 1218-1223.
De Berker, A. O., Rutledge, R. B., Mathys, C., Marshall, L., Cross, G. F., Dolan, R. J., & Bestmann, S. (2016). Computations of uncertainty mediate acute stress responses in humans. *Nature communications, 7*, 10996.
de Oliveira-Souza, R., Hare, R. D., Bramati, I. E., Garrido, G. J., Ignácio, F. A., Tovar-Moll, F., & Moll, J. (2008). Psychopathy as a disorder of the moral brain: fronto-temporo-limbic grey matter reductions demonstrated by voxel-based morphometry. *Neuroimage, 40*(3), 1202-1213.
Decety, J. (2010). To what extent is the experience of empathy mediated by shared neural circuits? *Emotion Review, 2*(3), 204-207.
Decety, J., Chen, C., Harenski, C., & Kiehl, K. A. (2013). An fMRI study of affective perspective taking in individuals with psychopathy: imagining another in pain does not evoke empathy. *Front Hum Neurosci, 7*, 489. doi:10.3389/fnhum.2013.00489
Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behav Cogn Neurosci Rev, 3*(2), 71-100. doi:10.1177/1534582304267187
Decety, J., Michalska, K. J., Akitsuki, Y., & Lahey, B. B. (2009). Atypical empathic responses in adolescents with aggressive conduct disorder: a functional MRI investigation. *Biol Psychol, 80*(2), 203-211. doi:10.1016/j.biopsycho.2008.09.004

Decety, J., & Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: implications for intervention across different clinical conditions. *Biopsychosoc Med, 1*, 22. doi:10.1186/1751-0759-1-22
Declercq, F., Carter, R., & Neumann, C. S. (2015). Assessing psychopathic traits and criminal behavior in a young adult female community sample using the Self- Report Psychopathy Scale. *Journal of forensic sciences, 60*(4), 928-935.
Decuyper, M., De Pauw, S., De Fruyt, F., De Bolle, M., & De Clercq, B. J. (2009). A meta- analysis of psychopathy- , antisocial PD- and FFM associations. *European Journal of Personality, 23*(7), 531-565.
Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological science, 5*(5), 303-305.
DeLisi, M. (2016). Psychopathy, Assessment, and the Structure of Crime. In *Psychopathy as Unified Theory of Crime* (pp. 61-95): Springer.
DeLisi, M., Beaver, K. M., Wright, K. A., Wright, J. P., Vaughn, M. G., & Trulson, C. R. (2011). Criminal Specialization Revisited: A Simultaneous Quantile Regression Approach. *American Journal of Criminal Justice, 36*(2), 73-92. doi:10.1007/s12103-010-9083-1
DeLisi, M., Tostlebe, J., Burgason, K., Heirigs, M., & Vaughn, M. G. (2016). Self-Control Versus Psychopathy: A Head-to-Head Test of General Theories of Antisociality. *Youth Violence and Juvenile Justice, 16*(1), 53-76. doi:10.1177/1541204016682998
DeLisi, M., Vaughn, M. G., Salas-Wright, C. P., & Jennings, W. G. (2015). Drugged and Dangerous: Prevalence and Variants of Substance Use Comorbidity Among Seriously Violent Offenders in the United States. *Journal of Drug Issues, 45*(3), 232-248. doi:10.1177/0022042615579237
DeMatteo, D., Heilbrun, K., & Marczyk, G. (2006). An empirical investigation of psychopathy in a noninstitutionalized and noncriminal sample. *Behavioral sciences & the law, 24*(2), 133-146.
Diaconescu, A. O., Mathys, C., Weber, L. A., Daunizeau, J., Kasper, L., Lomakina, E. I., . . . Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical Bayesian learning. *PLoS Comput Biol, 10*(9), e1003810.
Diaconescu, A. O., Mathys, C., Weber, L. A., Kasper, L., Mauer, J., & Stephan, K. E. (2017). Hierarchical prediction errors in midbrain and septum during social learning. *Social cognitive and affective neuroscience, 12*(4), 618-634.
Dolan, M. C., & Fullam, R. S. (2009). Psychopathy and functional magnetic resonance imaging blood oxygenation level-dependent responses to emotional faces in violent patients with schizophrenia. *Biological psychiatry, 66*(6), 570-577.
Driessen, J. M., Fanti, K. A., Glennon, J. C., Neumann, C. S., Baskin-Sommers, A. R., & Brazil, I. A. (2018). A comparison of latent profiles in antisocial male offenders. *Journal of Criminal Justice, 57*, 47-55.
Dufwenberg, M., & Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and economic Behavior, 30*(2), 163-182.

E

Eisenbarth, H., Angrilli, A., Calogero, A., Harper, J., Olson, L. A., & Bernat, E. (2013). Reduced negative affect response in female psychopaths. *Biological psychology, 94*(2), 310-318.
Elliott, F. A. (1978). Neurological aspects of antisocial behavior. *The psychopath: A comprehensive study of antisocial disorders and behaviors*, 146-189.
Engelmann, J. B., Schmid, B., De Dreu, C. K., Chumbley, J., & Fehr, E. (2019). On the psychology and economics of antisocial personality. *Proceedings of the National Academy of Sciences, 116*(26), 12781-12786.
Enter, D., Spinhoven, P., & Roelofs, K. (2014). Alleviating social avoidance: Effects of single dose testosterone administration on approach–avoidance action. *Hormones and behavior, 65*(4), 351-354.
Enter, D., Spinhoven, P., & Roelofs, K. (2016). Dare to approach: single dose testosterone administration promotes threat approach in patients with social anxiety disorder. *Clinical Psychological Science, 4*(6), 1073-1079.
Ermer, E., Cope, L. M., Nyalakanti, P. K., Calhoun, V. D., & Kiehl, K. A. (2012). Aberrant paralimbic gray matter in criminal psychopathy. *Journal of Abnormal Psychology, 121*(3), 649.

Ernst, M., & Fudge, J. L. (2009). A developmental neurobiological model of motivated behavior: anatomy, connectivity and ontogeny of the triadic nodes. *Neuroscience & Biobehavioral Reviews, 33*(3), 367-382.

Estévez, E., & Emler, N. P. (2011). Assessing the links among adolescent and youth offending, antisocial behaviour, victimization, drug use, and gender. *International Journal of Clinical and Health Psychology, 11*(2).

Etkin, A., Büchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nature reviews neuroscience, 16*(11), 693.

F

Falkenbach, D., Poythress, N., Falki, M., & Manchak, S. (2007). Reliability and Validity of Two Self-Report Measures of Psychopathy. *Assessment, 14*(4), 341-350. doi:10.1177/1073191107305612

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalogr Clin Neurophysiol, 78*(6), 447-455.

Fan, Y., & Han, S. (2008). Temporal dynamic of neural mechanisms involved in empathy for pain: an event-related brain potential study. *Neuropsychologia, 46*(1), 160-173.

Fang, C. Y., Egleston, B. L., Brown, K. M., Lavigne, J. V., Stevens, V. J., Barton, B. A., . . . Dorgan, J. F. (2009). Family cohesion moderates the relation between free testosterone and delinquent behaviors in adolescent boys and girls. *Journal of Adolescent Health, 44*(6), 590-597.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics, 114*(3), 817-868.

Feilhauer J., Cima, M. (2013). Differential correlates of callous-unemotional traits, narcissism, and impulsivity. *Forensic Science International, 224*(1-3), 1–7. 10;224(1-3):1-7. doi: 10.1016/j.forsciint.2012.10.016. Epub 2012 Nov 11. PMID: 23149319.

Ferdinand, N. K., Mecklinger, A., Kray, J., & Gehring, W. J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. *Journal of Neuroscience, 32*(35), 12087-12092.

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., & Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature neuroscience, 22*(12), 2066-2077.

Fowles, D. C. (1980). The three arousal model: Implications of Gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology, 17*(2), 87-104.

Fowles, D. C. (1988). Psychophysiology and psychopathology: A motivational approach. *Psychophysiology, 25*(4), 373-391.

Franken, I. H., Muris, P., & Rassin, E. (2005). Psychometric properties of the Dutch BIS/BAS scales. *Journal of Psychopathology and Behavioral Assessment, 27*(1), 25-30.

G

Gao, Y., & Raine, A. (2010). Successful and unsuccessful psychopaths: A neurobiological model. *Behavioral sciences & the law, 28*(2), 194-210.

Gatzke-Kopp, L. M., Beauchaine, T. P., Shannon, K. E., Chipman, J., Fleming, A. P., Crowell, S. E., . . . Aylward, E. (2009). Neurological correlates of reward responding in adolescents with and without externalizing behavior disorders. *Journal of Abnormal Psychology, 118*(1), 203.

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological science, 4*(6), 385-390.

Gehring, W. J., Himle, J., & Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychological science, 11*(1), 1-6.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*(5563), 2279-2282.

Glenn, A. L., Iyer, R., Graham, J., Koleva, S., & Haidt, J. (2009). Are all types of morality compromised in psychopathy? *Journal of personality disorders, 23*(4), 384-398.

Glenn, A. L., Raine, A., & Schug, R. A. (2009). The neural correlates of moral decision-making in psychopathy. *Molecular psychiatry, 14*(1), 5-6.

Glenn, A. L., Raine, A., Schug, R. A., Gao, Y., & Granger, D. A. (2011). Increased testosterone-to-cortisol ratio in psychopathy. *Journal of abnormal psychology, 120*(2), 389.

Glenn, A. L., Raine, A., Yaralian, P. S., & Yang, Y. (2010). Increased volume of the striatum in psychopathic individuals. *Biological psychiatry, 67*(1), 52-58.

Glenn, A. L., & Yang, Y. (2012). The potential role of the striatum in antisocial behavior and psychopathy. *Biological psychiatry, 72*(10), 817-822.

Gong, X., Brazil, I. A., Chang, L. J., & Sanfey, A. G. (2019). Psychopathic traits are related to diminished guilt aversion and reduced trustworthiness during social decision-making. *Scientific reports, 9*(1), 1-11.

Gordon, H. L., Baird, A. A., & End, A. (2004). Functional differences among those high and low on a trait measure of psychopathy. *Biological psychiatry, 56*(7), 516-521.

Gordts, S., Uzieblo, K., Neumann, C., Van den Bussche, E., & Rossi, G. (2017). Validity of the Self-Report Psychopathy Scales (SRP-III full and short versions) in a community sample. *Assessment, 24*(3), 308-325.

Gorenstein, E. E. (1982). Frontal lobe functions in psychopaths. *Journal of Abnormal Psychology, 91*(5), 368.

Gorenstein, E. E., & Newman, J. P. (1980). Disinhibitory psychopathology: A new perspective and a model for research. *Psychological review, 87*(3), 301.

Gray, H. M., Ambady, N., Lowenthal, W. T., & Deldin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of experimental social psychology, 40*(2), 216-224.

Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. *Behaviour research and therapy, 8*(3), 249-266.

Gregory, S., Blair, R. J., Simmons, A., Kumari, V., Hodgins, S., & Blackwood, N. (2015). Punishment and psychopathy: a case-control functional MRI investigation of reinforcement learning in violent antisocial personality disordered men. *The Lancet Psychiatry, 2*(2), 153-160.

H

Hall, J. R., & Benning, S. D. (2006). The "successful" psychopath. *Handbook of psychopathy*, 459-478.

Hansen, A. L., Johnsen, B. H., Thornton, D., Waage, L., & Thayer, J. F. (2007). Facets of psychopathy, heart rate variability and cognitive function. *Journal of personality disorders, 21*(5), 568-582.

Hare, R. D. (1965). Psychopathy, fear arousal and anticipated pain. *Psychological Reports, 16*(2), 499-502.

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and individual differences, 1*(2), 111-119.

Hare, R. D. (1985). Comparison of procedures for the assessment of psychopathy. *Journal of consulting and clinical psychology, 53*(1), 7.

Hare, R. D. (1991). *The Hare psychopathy checklist-revised: Manual*: Multi-Health Systems, Incorporated.

Hare, R. D. (1999). *Without conscience: The disturbing world of the psychopaths among us*: Guilford Press.

Hare, R. D. (2003). The psychopathy checklist–Revised. *Toronto, ON.*

Hare, R. D. (2016). Psychopathy, the PCL-R, and criminal justice: Some new findings and current issues. *Canadian Psychology/Psychologie canadienne, 57*(1), 21-34. doi:10.1037/cap0000041

Hare, R. D., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: An international perspective. *Behavioral sciences & the law, 18*(5), 623-645.

Hare, R. D., Forth, A. E., & Hart, S. D. (1989). The psychopath as prototype for pathological lying and deception. In Credibility assessment (pp. 25-49). Springer, Dordrecht.

Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P. (1990). The revised Psychopathy Checklist: Reliability and factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*(3), 338.

Hare, R. D., & Neumann, C. S. (2005). Structural models of psychopathy. *Current psychiatry reports, 7*(1), 57-64.

Hare, R. D., & Neumann, C. S. (2006). The PCL-R assessment of psychopathy. *Handbook of psychopathy*, 58-88.

Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annu. Rev. Clin. Psychol., 4*, 217-246.

Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology, 4*, 217-246.

Hare, R. D., & Neumann, C. S. (2009). Psychopathy: Assessment and forensic implications. *The Canadian Journal of Psychiatry, 54*(12), 791-802.
Hare, R. D., & Neumann, C. S. (2010). The role of antisociality in the psychopathy construct: Comment on Skeem and Cooke (2010).
Harenski, C. L., Harenski, K. A., & Kiehl, K. A. (2014). Neural processing of moral violations among incarcerated adolescents with psychopathic traits. *Developmental cognitive neuroscience, 10*, 181-189.
Harenski, C. L., Harenski, K. A., Shane, M. S., & Kiehl, K. A. (2010). Aberrant neural processing of moral violations in criminal psychopaths. *Journal of Abnormal Psychology, 119*(4), 863.
Harenski, C. L., Kim, S. H., & Hamann, S. (2009). Neuroticism and psychopathy predict brain activation during moral and nonmoral emotion regulation. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 1-15.
Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., & Weinberger, D. R. (2002). The amygdala response to emotional stimuli: a comparison of faces and scenes. *Neuroimage, 17*(1), 317-323.
Harpur, T. J., Hakstian, A. R., & Hare, R. D. (1988). Factor structure of the Psychopathy Checklist. *Journal of consulting and clinical psychology, 56*(5), 741.
Harris, G. T., & Rice, M. E. (2006). Treatment of psychopathy. *Handbook of psychopathy*, 555-572.
Hastings, M. E., Tangney, J. P., & Stuewig, J. (2008). Psychopathy and identification of facial expressions of emotion. *Personality and individual differences, 44*(7), 1474-1483.
Hauser, T. U., Iannaccone, R., Ball, J., Mathys, C., Brandeis, D., Walitza, S., & Brem, S. (2014). Role of the medial prefrontal cortex in impaired decision making in juvenile attention-deficit/hyperactivity disorder. *JAMA psychiatry, 71*(10), 1165-1173.
Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication monographs, 76*(4), 408-420.
Hein, G., Morishima, Y., Leiberg, S., Sul, S., & Fehr, E. (2016). The brain's functional network architecture reveals human motives. *science, 351*(6277), 1074-1078.
Hermans, E. J., Ramsey, N. F., & van Honk, J. (2008). Exogenous testosterone enhances responsiveness to social threat in the neural circuitry of social aggression in humans. *Biological psychiatry, 63*(3), 263-270.
Heuer, K., Rinck, M., & Becker, E. S. (2007). Avoidance of emotional facial expressions in social anxiety: The approach–avoidance task. *Behaviour research and therapy, 45*(12), 2990-3001.
Hicks, B. M., Markon, K. E., Patrick, C. J., Krueger, R. F., & Newman, J. P. (2004). Identifying psychopathy subtypes on the basis of personality structure. *Psychological assessment, 16*(3), 276.
Hicks, B. M., & Patrick, C. J. (2006). Psychopathy and Negative Emotionality: Analyses of Suppressor Effects Reveal Distinct Relations With Emotional Distress, Fearfulness, and Anger–Hostility. *Journal of abnormal psychology, 115*(2), 276-287. doi:10.1037/0021-843X.115.2.276
Hill, C. D., Neumann, C. S., & Rogers, R. (2004). Confirmatory factor analysis of the psychopathy checklist: screening version in offenders with axis I disorders. *Psychological Assessment, 16*(1), 90.
Hindelang, M. J., Hirschi, T., & Weis, J. G. (1981). *Measuring delinquency*: Sage Publications Beverly Hills.
Hoppenbrouwers, S. S., Bulten, B. H., & Brazil, I. A. (2016). Parsing fear: A reassessment of the evidence for fear deficits in psychopathy. *Psychological Bulletin, 142*(6), 573.
Hoppenbrouwers, S. S., Neumann, C. S., Lewis, J., & Johansson, P. (2015). A latent variable analysis of the Psychopathy Checklist–Revised and behavioral inhibition system/behavioral activation system factors in North American and Swedish offenders. *Personality Disorders: Theory, Research, and Treatment, 6*(3), 251.
Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P., & Polkey, C. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain, 126*(7), 1691-1712.
Horvitz, J. C. (2009). Stimulus–response and response–outcome learning mechanisms in the striatum. *Behavioural brain research, 199*(1), 129-140.

I

Iannetti, G., Zambreanu, L., Cruccu, G., & Tracey, I. (2005). Operculoinsular cortex encodes pain intensity at the earliest stages of cortical processing as indicated by amplitude of laser-evoked potentials in humans. *Neuroscience, 131*(1), 199-208.
Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E., & Stephan, K. E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron, 80*(2), 519-530.

J

Jackson, R. L., Neumann, C. S., & Vitacco, M. J. (2007). Impulsivity, anger, and psychopathy: The moderating effect of ethnicity. *Journal of personality disorders, 21*(3), 289-304.
Jensen, P. S., Hinshaw, S. P., Kraemer, H. C., Lenora, N., Newcorn, J. H., Abikoff, H. B., . . . Conners, C. K. (2001). ADHD comorbidity findings from the MTA study: comparing comorbid subgroups. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(2), 147-158.
Jocham, G., Neumann, J., Klein, T. A., Danielmeier, C., & Ullsperger, M. (2009). Adaptive coding of action values in the human rostral cingulate zone. *Journal of Neuroscience, 29*(23), 7489-7496.
Johanson, M., Vaurio, O., Tiihonen, J., & Lähteenvuo, M. (2020). A systematic literature review of neuroimaging of psychopathic traits. *Frontiers in psychiatry, 10*, 1027.
Justus, A. N., & Finn, P. R. (2007). Startle modulation in non-incarcerated men and women with psychopathic traits. *Personality and individual differences, 43*(8), 2057-2071.

K

Kaldewaij, R., Koch, S. B., Volman, I., Toni, I., & Roelofs, K. (2016). On the control of social approach–avoidance behavior: neural and endocrine mechanisms. In *Social Behavior from Rodents to Humans* (pp. 275-293): Springer.
Kandel, E., Mednick, S. A., Kirkegaard-Sorensen, L., Hutchings, B., Knop, J., Rosenberg, R., & Schulsinger, F. (1988). IQ as a protective factor for subjects at high risk for antisocial behavior. *Journal of Consulting and Clinical psychology, 56*(2), 224.
Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of science, 78*(4), 601-627.
Katsarava, Z., Ayzenberg, I., Sack, F., Limmroth, V., Diener, H. C., & Kaube, H. (2006). A Novel Method of Eliciting Pain-Related Potentials by Transcutaneous Electrical Stimulation. *Headache: The Journal of Head and Face Pain, 46*(10), 1511-1517.
Kaya, S., Yildirim, H., & Atmaca, M. (2020). Reduced hippocampus and amygdala volumes in antisocial personality disorder. *Journal of Clinical Neuroscience, 75*, 199-203.
Khalmetski, K. (2016). Testing guilt aversion with an exogenous shift in beliefs. *Games and economic Behavior, 97*, 110-119.
Kiehl, K. A., Hare, R. D., McDonald, J. J., & Brink, J. (1999). Semantic and affective processing in psychopaths: An event-related potential (ERP) study. *Psychophysiology, 36*(6), 765-774.
Kiehl, K. A., & Hoffman, M. B. (2011). The criminal psychopath: History, neuroscience, treatment, and economics. *Jurimetrics, 51*, 355.
Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., & Liddle, P. F. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological psychiatry, 50*(9), 677-684.
Kim, S., & Lee, D. (2011). Prefrontal cortex and impulsive decision making. *Biological psychiatry, 69*(12), 1140-1146.
Kimonis, E. R., Fanti, K. A., Isoma, Z., & Donoghue, K. (2013). Maltreatment profiles among incarcerated boys with callous-unemotional traits. *Child maltreatment, 18*(2), 108-121.
Kimonis, E. R., Goulter, N., Hawes, D. J., Wilbur, R. R., & Groer, M. W. (2017). Neuroendocrine factors distinguish juvenile psychopathy variants. *Developmental psychobiology, 59*(2), 161-173.
King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *science, 308*(5718), 78-83.
Klein, A. M., Becker, E. S., & Rinck, M. (2011). Approach and avoidance tendencies in spider fearful children: The approach-avoidance task. *Journal of Child and Family Studies, 20*(2), 224-231.

Koenigs, M., Kruepke, M., & Newman, J. P. (2010). Economic decision-making in psychopathy: a comparison with ventromedial prefrontal lesion patients. *Neuropsychologia, 48*(7), 2198-2204.

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *Journal of Neuroscience, 27*(4), 951-956.

Korponay, C., Pujara, M., Deming, P., Philippi, C., Decety, J., Kosson, D. S.,...Koenigs, M. (2017). Impulsive-antisocial dimension of psychopathy linked to enlargement and abnormal functional connectivity of the striatum. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2, 149–157.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435*(7042), 673-676.

Kosson, D. S., Cyterski, T. D., Steuerwald, B. L., Neumann, C. S., & Walker-Matthews, S. (2002). The reliability and validity of the Psychopathy Checklist: Youth Version (PCL: YV) in nonincarcerated adolescent males. *Psychological Assessment, 14*(1), 97.

Krekels, G., & Pandelaere, M. (2015). Dispositional greed. Personality and Individual Differences, 74, 225-230.

Krstic, S., Neumann, C. S., Roy, S., Robertson, C. A., Knight, R. A., & Hare, R. D. (2018). Using latent variable-and person-centered approaches to examine the role of psychopathic traits in sex offenders. *Personality Disorders: Theory, Research, and Treatment, 9*(3), 207.

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking Antisocial Behavior, Substance Use, and Personality: An Integrative Quantitative Model of the Adult Externalizing Spectrum. *Journal of abnormal psychology, 116*(4), 645-666. doi:10.1037/0021-843X.116.4.645

Krugel, L. K., Biele, G., Mohr, P. N., Li, S.-C., & Heekeren, H. R. (2009). Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proceedings of the National Academy of Sciences, 106*(42), 17951-17956.

Kuba, M., Kremláček, J., Langrová, J., Kubová, Z., Szanyi, J., & Vít, F. (2012). Aging effect in pattern, motion and cognitive visual evoked potentials. *Vision research, 62*, 9-16.

L

Lang, P. J., & Bradley, M. M. (2010). Emotion and the motivational brain. *Biological psychology, 84*(3), 437-450.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. *Attention and orienting: Sensory and motivational processes, 97*, 135.

Lange, W.-G., Keijsers, G., Becker, E. S., & Rinck, M. (2008). Social anxiety and evaluation of social crowds: Explicit and implicit measures. *Behaviour research and therapy, 46*(8), 932-943.

Lange, W.-G., Salemink, E., Windey, I., Keijsers, G. P., Krans, J., Becker, E. S., & Rinck, M. (2010). Does modified interpretation bias influence automatic avoidance behaviour? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 24*(3), 326-337.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and emotion, 24*(8), 1377-1388.

Larson, M. J., Fair, J. E., Good, D. A., & Baldwin, S. A. (2010). Empathy and error processing. *Psychophysiology, 47*(3), 415-424.

Laurell, J., Belfrage, H., & Hellström, Å. (2010). Facets on the psychopathy checklist screening version and instrumental violence in forensic psychiatric patients. *Criminal Behaviour and Mental Health, 20*(4), 285-294. doi:10.1002/cbm.779

Lee, Z., Salekin, R. T., & Iselin, A.-M. R. (2010). Psychopathic traits in youth: is there evidence for primary and secondary subtypes? *Journal of abnormal child psychology, 38*(3), 381-393.

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of personality and social psychology, 68*(1), 151.

Lietz, C. A., Gerdes, K. E., Sun, F., Geiger, J. M., Wagaman, M. A., & Segal, E. A. (2011). The Empathy Assessment Index (EAI): A confirmatory factor analysis of a multidimensional model of empathy. *Journal of the Society for Social Work and Research, 2*(2), 104-124.

Lilienfed, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in non-criminal populations. *Journal of Personality Assessment, 66*(3), 448-524.

Lilienfeld, S. O., & Widows, M. (2005). Professional manual for the psychopathic personality inventory-revised (PPI-R). *Lutz, FL: Psychological Assessment Resources.*

Lilienfeld, S. O., Fowler, K. A., & Patrick, C. (2006). The self-report assessment of psychopathy. *Handbook of psychopathy*, 107-132.

Lombardo, M. V., Ashwin, E., Auyeung, B., Chakrabarti, B., Lai, M.-C., Taylor, K., . . . Baron-Cohen, S. (2012). Fetal programming effects of testosterone on the reward system and behavioral approach tendencies in humans. *Biological psychiatry, 72*(10), 839-847.

Ly, V., von Borries, A. K. L., Brazil, I. A., Bulten, B. H., Cools, R., & Roelofs, K. (2016). Reduced transfer of affective value to instrumental behavior in violent offenders. *Journal of Abnormal Psychology, 125*(5), 657.

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *The Journal of Abnormal and Social Psychology, 55*(1), 6.

Lykken, D. T. (1995). *The antisocial personalities*: Psychology Press.

Lynam, D. R., Caspi, A., Moffitt, T. E., Loeber, R., & Stouthamer-Loeber, M. (2007). Longitudinal evidence that psychopathy scores in early adolescence predict adult psychopathy. *Journal of abnormal psychology, 116*(1), 155.

M

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature neuroscience, 14*(2), 154-162.

Malterer, M. B., Lilienfeld, S. O., Neumann, C. S., & Newman, J. P. (2010). Concurrent validity of the Psychopathic Personality Inventory with offender and community samples. *Assessment, 17*(1), 3-15.

Maras, A., Laucht, M., Gerdes, D., Wilhelm, C., Lewicka, S., Haack, D., . . . Schmidt, M. H. (2003). Association of testosterone and dihydrotestosterone with externalizing behavior in adolescent boys and girls. *Psychoneuroendocrinology, 28*(7), 932-940.

Marco-Pallares, J., Cucurell, D., Cunillera, T., García, R., Andrés-Pueyo, A., Münte, T. F., & Rodríguez-Fornells, A. (2008). Human oscillatory activity associated to reward processing in a gambling task. *Neuropsychologia, 46*(1), 241-248.

Marcoux, L.-A., Michon, P.-E., Lemelin, S., Voisin, J. A., Vachon-Presseau, E., & Jackson, P. L. (2014). Feeling but not caring: empathic alteration in narcissistic men with high psychopathic traits. *Psychiatry Research: Neuroimaging, 224*(3), 341-348.

Mars, R. B., Shea, N. J., Kolling, N., & Rushworth, M. F. (2012). Model-based analyses: promises, pitfalls, and example applications to the study of cognitive control. *Quarterly journal of experimental psychology, 65*(2), 252-267.

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach-and avoidance-related behaviors. *Emotion, 5*(1), 119.

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.

Marsh, A. A., & Cardinale, E. M. (2014). When psychopathy impairs moral judgments: neural responses during judgments about causing fear. *Social cognitive and affective neuroscience, 9*(1), 3-11.

Marsh, A. A., Finger, E. C., Fowler, K. A., Jurkowitz, I. T., Schechter, J. C., Henry, H. Y., . . . Blair, R. J. R. (2011). Reduced amygdala–orbitofrontal connectivity during moral judgments in youths with disruptive behavior disorders and psychopathic traits. *Psychiatry Research: Neuroimaging, 194*(3), 279-286.

Marsh, A. A., Stoycos, S. A., Brethel-Haurwitz, K. M., Robinson, P., VanMeter, J. W., & Cardinale, E. M. (2014). Neural and cognitive characteristics of extraordinary altruists. *Proceedings of the National Academy of Sciences, 111*(42), 15036-15041.

Marshall, L., Mathys, C., Ruge, D., De Berker, A. O., Dayan, P., Stephan, K. E., & Bestmann, S. (2016). Pharmacological fingerprints of contextual uncertainty. *PLoS biology, 14*(11), e1002575.

Marquand, A.F., Kia, S.M., Zabihi, M., Wolfers, T., Buitelaar, J. K., Beckmann, C. F. (2019). Conceptualizing mental disorders as deviations from normative functioning. *Mol Psychiatry, 24,* 1415–1424. https://doi.org/10.1038/s41380-019-0441-1

Mas-Herrero, E., & Marco-Pallarés, J. (2014). Frontal theta oscillatory activity is a common mechanism for the computation of unexpected outcomes and learning rate. *Journal of cognitive neuroscience, 26*(3), 447-458.
Mathewson, K. E., Harrison, T. J., & Kizuk, S. A. (2017). High and dry? Comparing active dry EEG electrodes to active and passive wet electrodes. *Psychophysiology, 54*(1), 74-82.
Mathieu, C., & Babiak, P. (2016). Corporate psychopathy and abusive supervision: Their influence on employees' job satisfaction and turnover intentions. *Personality and Individual Differences, 91*, 102-106.
Mathys, C. D. (2012). *Hierarchical Gaussian filtering: Construction and variational inversion of a generic Bayesian model of individual learning under uncertainty.* ETH Zurich,
Mathys, C. D., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in human neuroscience, 5*, 39.
Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in human neuroscience, 8*, 825.
Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nature neuroscience, 10*(5), 647-656.
McCuish, E. C., Corrado, R. R., Hart, S. D., & DeLisi, M. (2015). The role of symptoms of psychopathy in persistent violence over the criminal career into full adulthood. *Journal of Criminal Justice, 43*(4), 345-356. doi:https://doi.org/10.1016/j.jcrimjus.2015.04.008
Mednick, S. A., Vka, J. V., Gabrielli, J., William F., & Itil, T. M. (1981). EEG as a predictor of antisocial behavior. *Criminology, 19*(2), 219-230.
Mehta, P. H., & Beer, J. (2010). Neural mechanisms of the testosterone–aggression relation: the role of orbitofrontal cortex. *Journal of cognitive neuroscience, 22*(10), 2357-2368.
Melzack, R., & Wall, P. D. (1967). Pain mechanisms: a new theory. *Survey of Anesthesiology, 11*(2), 89-90.
Meng, J., Hu, L., Shen, L., Yang, Z., Chen, H., Huang, X., & Jackson, T. (2012). Emotional primes modulate the responses to others' pain: an ERP study. *Experimental brain research, 220*(3-4), 277-286.
Meng, J., Jackson, T., Chen, H., Hu, L., Yang, Z., Su, Y., & Huang, X. (2013). Pain perception in the self and observation of others: an ERP investigation. *Neuroimage, 72*, 164-173.
Meng, W., Butterworth, J., Malecaze, F., & Calvas, P. (2012). A linear regression model to differentiate axial myopia from non-axial myopia. *Ophthalmic genetics, 33*(1), 53-55.
Milgram, S. (1963). Behavioral Study of obedience. *The Journal of abnormal and social psychology, 67*(4), 371.
Miller, J. D., & Lynam, D. R. (2015). Psychopathy and personality: Advances and debates. *Journal of personality, 83*(6), 585-592.
Millon, T., Lerner, M. J., & Weiner, I. B. (2003). *Handbook of psychology: Personality and social psychology* (Vol. 688): Wiley Hoboken, NJ.
Mitchell, D. G., Richell, R. A., Leonard, A., & Blair, R. J. R. (2006). Emotion at the expense of cognition: Psychopathic individuals outperform controls on an operant response task. *Journal of Abnormal Psychology, 115*(3), 559.
Moffitt, T. E. (1993). The neuropsychology of conduct disorder. *Development and psychopathology, 5*(1-2), 135-151.
Mokros, A., Hare, R. D., Neumann, C. S., Santtila, P., Habermeyer, E., & Nitschke, J. (2015a). Variants of psychopathy in adult male offenders: A latent profile analysis. *Journal of Abnormal Psychology, 124*(2), 372.
Mokros, A., Menner, B., Eisenbarth, H., Alpers, G. W., Lange, K. W., & Osterheider, M. (2008). Diminished cooperativeness of psychopaths in a prisoner's dilemma game yields higher rewards. *Journal of Abnormal Psychology, 117*(2), 406.
Moltó, J., Poy, R., Segarra, P., Pastor, M. C., & Montañés, S. (2007). Response perseveration in psychopaths: Interpersonal/affective or social deviance traits? *Journal of Abnormal Psychology, 116*(3), 632.
Morey, R. D., Rouder, J. N., Verhagen, J., & Wagenmakers, E.-J. (2014). Why hypothesis tests are essential for psychological science a comment on Cumming (2014). *Psychological science, 25*(6), 1289-1290.
Morey, R. D., & Wagenmakers, E.-J. (2014). Simple relation between Bayesian order-restricted and point-null hypothesis tests. *Statistics & Probability Letters, 92*, 121-124.

REFERENCES

Morgan, A. B., & Lilienfeld, S. O. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clinical Psychology Review, 20*(1), 113-136. doi:https://doi.org/10.1016/S0272-7358(98)00096-8
Müller, J. L., Sommer, M., Döhnel, K., Weber, T., Schmidt- Wilcke, T., & Hajak, G. (2008). Disturbed prefrontal and temporal brain function during emotion and cognition interaction in criminal psychopathy. *Behavioral sciences & the law, 26*(1), 131-150.
Müller, J. L., Sommer, M., Wagner, V., Lange, K., Taschler, H., Röder, C. H., . . . Hajak, G. (2003). Abnormalities in emotion processing within cortical and subcortical regions in criminal psychopaths: evidence from a functional magnetic resonance imaging study using pictures with emotional content. *Biological psychiatry, 54*(2), 152-162.
Mullis, R. J., Holcomb, P. J., Diner, B. C., & Dykman, R. A. (1985). The effects of aging on the P3 component of the visual event-related potential. *Electroencephalogr Clin Neurophysiol, 62*(2), 141-149.
Murphy, C., & Vess, J. (2003). Subtypes of psychopathy: Proposed differences between narcissistic, borderline, sadistic, and antisocial psychopaths. *Psychiatric quarterly, 74*(1), 11-29.
Muthén, B. (2000). Methodological issues in random coefficient growth modeling using a latent variable framework: Applications to the development of heavy drinking. *Multivariate applications in substance use research*, 113-140.

N

Nagin, D. (2005). *Group-based modeling of development*: Harvard University Press.
Neumann, C. S., & Hare, R. D. (2008). Psychopathic traits in a large community sample: links to violence, alcohol use, and intelligence. *Journal of consulting and clinical psychology, 76*(5), 893.
Neumann, C. S., Hare, R. D., & Newman, J. P. (2007). The super-ordinate nature of the Psychopathy Checklist-Revised. *Journal of personality disorders, 21*(2), 102-117.
Neumann, C. S., Johansson, P. T., & Hare, R. D. (2013). The Psychopathy Checklist-Revised (PCL-R), low anxiety, and fearlessness: a structural equation modeling analysis. *Personality Disorders, 4*(2), 129-137. doi:10.1037/a0027886
Neumann, C. S., & Pardini, D. (2014). Factor structure and construct validity of the Self-Report Psychopathy (SRP) Scale and the Youth Psychopathic Traits Inventory (YPI) in young men. *Journal of personality disorders, 28*(3), 419-433.
Neumann, C. S., & Pardini, D. (2014b). Factor structure and construct validity of the Self-Report Psychopathy (SRP) scale and the Youth Psychopathic Traits Inventory (YPI) in young men. *Journal of Personality Disorders, 28*(3), 419-433. doi:10.1521/pedi_2012_26_063
Neumann, C. S., Schmitt, D. S., Carter, R., Embley, I., & Hare, R. D. (2012). Psychopathic traits in females and males across the globe. *Behavioral Sciences and the Law, 30*(5), 557-574. doi:10.1002/bsl.2038
Neumann, C. S., Vitacco, M. J., Hare, R. D., & Wupperman, P. (2005). Reconstruing the "reconstruction" of psychopathy: A comment on Cooke, Michie, Hart, and Clark. *Journal of personality disorders, 19*(6), 624-640.
Neumann, C. S., Vitacco, M. J., & Mokros, A. S. (2016). Using both variable-centered and person-centered approaches to understanding psychopathic personality. *The clinical and forensic assessment of psychopathy: A practitioner's guide*, 14-31.
Newman, J. P. (1998). Psychopathic behavior: An information processing perspective. In *Psychopathy: Theory, research and implications for society* (pp. 81-104): Springer.
Newman, J. P., & Kosson, D. S. (1986). Passive avoidance learning in psychopathic and nonpsychopathic offenders. *Journal of Abnormal Psychology, 95*(3), 252.
Newman, J. P., MacCoon, D. G., Vaughn, L. J., & Sadeh, N. (2005). Validating a distinction between primary and secondary psychopathy with measures of Gray's BIS and BAS constructs. *Journal of Abnormal Psychology, 114*(2), 319.
Nihonsugi, T., Ihara, A., & Haruno, M. (2015). Selective increase of intention-based economic decisions by non-invasive brain stimulation to the dorsolateral prefrontal cortex. *Journal of Neuroscience, 35*(8), 3412-3419.
Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535-569. doi:10.1080/10705510701575396

O

Oba, T., Katahira, K., & Ohira, H. (2019). The Effect of Reduced Learning Ability on Avoidance in Psychopathy: A Computational Approach. *Frontiers in psychology, 10.*

Odgers, C. L., Moretti, M. M., Burnette, M. L., Chauhan, P., Waite, D., & Reppucci, N. D. (2007). A latent variable modeling approach to identifying subtypes of serious and violent female juvenile offenders. *Aggressive Behavior, 33*(4), 339-352. doi:10.1002/ab.20190

Oldehinkel, M., Beckmann, C. F., Pruim, R. H., Van Oort, E. S., Franke, B., Hartman, C. A., ... & Mennes, M. (2016). Attention-deficit/hyperactivity disorder symptoms coincide with altered striatal connectivity. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1(4)*, 353-363.

Olver, M. E., Neumann, C. S., Wong, S. C., & Hare, R. D. (2013). The structural and predictive properties of the Psychopathy Checklist–Revised in Canadian Aboriginal and non-Aboriginal offenders. *Psychological Assessment, 25*(1), 167.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011.*

Osumi, T., & Ohira, H. (2010). The positive side of psychopathy: Emotional detachment in psychopathy and rational decision-making in the ultimatum game. *Personality and Individual Differences, 49*(5), 451-456.

Osumi, T., & Ohira, H. (2017). Selective fair behavior as a function of psychopathic traits in a subclinical population. *Frontiers in psychology, 8*, 1604.

P

Patrick, C. J. (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology, 31*(4), 319-330.

Patrick, C. J. (2018). Cognitive and emotional processing in psychopathy. *Handbook of psychopathy, 2.*

Patrick, C. J., & Bernat, E. M. (2009). Neurobiology of psychopathy: A two process theory.

Patrick, C. J., Cuthbert, B. N., & Lang, P. J. (1994). Emotion in the criminal psychopath: fear image processing. *Journal of abnormal psychology, 103*(3), 523.

Patrick, C.J., Fowles, D.C., Krueger, R.F (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and psychopathology, 21*(3), 913– 938.

Patrick, C. J., Hicks, B. M., Krueger, R. F., & Lang, A. R. (2005). Relations between psychopathy facets and externalizing in a criminal offender sample. *Journal of personality disorders, 19*(4), 339-356.

Patterson, C. M., & Newman, J. P. (1993). Reflectivity and learning from aversive events: Toward a psychological mechanism for the syndromes of disinhibition. *Psychological review, 100*(4), 716.

Paulhus, D., Neumann, C. S., & Hare, R. D. (2015). Self-report psychopathy scale: version III. Toronto. *ON: Multi-Health Systems.*

Paulhus, D. L., Neumann, C. S., & Hare, R. D. (2009). Manual for the self-report psychopathy scale. *Toronto: Multi-health systems.*

Paulus, M. P., Huys, Q. J., & Maia, T. V. (2016). A roadmap for the development of applied computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1*(5), 386-392.

Pellino, T. A., & Ward, S. E. (1998). Perceived control mediates the relationship between pain severity and patient satisfaction. *J Pain Symptom Manage, 15*(2), 110-116.

Peper, J. S., van den Heuvel, M. P., Mandl, R. C., Pol, H. E. H., & van Honk, J. (2011). Sex steroids and connectivity in the human brain: a review of neuroimaging studies. *Psychoneuroendocrinology, 36*(8), 1101-1113.

Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R., . . . Rugg, M. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology, 37*(02), 127-152.

Popma, A., Vermeiren, R., Geluk, C. A., Rinne, T., van den Brink, W., Knol, D. L., . . . Doreleijers, T. A. (2007). Cortisol moderates the relationship between testosterone and aggression in delinquent male adolescents. *Biological psychiatry, 61*(3), 405-411.

Porter, S., Birt, A. R., & Boer, D. P. (2001). Investigation of the Criminal and Conditional Release Profiles of Canadian Federal Offenders as a Function of Psychopathy and Age. *Law and Human Behavior, 25*(6), 647-661.

Poythress, N. G., Skeem, J. L., & Lilienfeld, S. O. (2006). Associations among early abuse, dissociation, and psychopathy in an offender sample. *Journal of Abnormal Psychology, 115*(2), 288.

Prado, C. E., Treeby, M. S., & Crowe, S. F. (2016). Examining the relationships between sub-clinical psychopathic traits with shame, guilt and externalisation response tendencies to everyday transgressions. *The Journal of Forensic Psychiatry & Psychology, 27*(4), 569-585.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology, 67*(4), 741.

Puca, R. M., Rinkenauer, G., & Breidenstein, C. (2006). Individual differences in approach and avoidance movements: How the avoidance motive influences response force. *Journal of Personality, 74*(4), 979-1014.

Pujol, J., Batalla, I., Contreras-Rodríguez, O., Harrison, B. J., Pera, V., Hernández-Ribas, R., . . . Deus, J. (2012). Breakdown in the brain network subserving moral judgment in criminal psychopathy. *Social cognitive and affective neuroscience, 7*(8), 917-923.

R

Radke, S., Seidel, E.-M., Eickhoff, S. B., Gur, R. C., Schneider, F., Habel, U., & Derntl, B. (2016). When opportunity meets motivation: Neural engagement during social approach is linked to high approach motivation. *Neuroimage, 127*, 267-276.

Radke, S., Volman, I., Mehta, P., van Son, V., Enter, D., Sanfey, A., . . . Roelofs, K. (2015). Testosterone biases the amygdala toward social threat approach. *Science advances, 1*(5), e1400074.

Raine, A. (2002). Biosocial studies of antisocial and violent behavior in children and adults: A review. *Journal of abnormal child psychology, 30*(4), 311-326.

Raine, A., Lencz, T., Bihrle, S., LaCasse, L., & Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Archives of general psychiatry, 57*(2), 119-127.

Reidy, D. E., Shelley-Tremblay, J. F., & Lilienfeld, S. O. (2011). Psychopathy, reactive aggression, and precarious proclamations: A review of behavioral, cognitive, and biological research. *Aggression and Violent Behavior, 16*(6), 512-524.

Rilling, J. K., Glenn, A. L., Jairam, M. R., Pagnoni, G., Goldsmith, D. R., Elfenbein, H. A., & Lilienfeld, S. O. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biological psychiatry, 61*(11), 1260-1271.

Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *Journal of behavior therapy and experimental psychiatry, 38*(2), 105-120.

Roelofs, K. (2017). Freeze for action: neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1718), 20160206.

Roelofs, K., Minelli, A., Mars, R. B., Van Peer, J., & Toni, I. (2008). On the neural control of social emotional behavior. *Social Cognitive and Affective Neuroscience, 4*(1), 50-58.

Roelofs, K., Putman, P., Schouten, S., Lange, W.-G., Volman, I., & Rinck, M. (2010). Gaze direction differentially affects avoidance tendencies to happy and angry faces in socially anxious individuals. *Behaviour research and therapy, 48*(4), 290-294.

Ross, S. R., Moltó, J., Poy, R., Segarra, P., Pastor, M. C., & Montañés, S. (2007). Gray's model and psychopathy: BIS but not BAS differentiates primary from secondary psychopathy in noninstitutionalized young adults. *Personality and individual differences, 43*(7), 1644-1655. doi:https://doi.org/10.1016/j.paid.2007.04.020

Ross, S. R., & Rausch, M. K. (2001). Psychopathic attributes and achievement dispositions in a college sample. *Personality and Individual Differences, 30*(3), 471-480.

Rotteveel, M., & Phaf, R. H. (2004). Automatic affective evaluation does not automatically predispose for arm flexion and extension. *Emotion, 4*(2), 156.

Rudebeck, P. H., Buckley, M. J., Walton, M. E., & Rushworth, M. F. (2006). A role for the macaque anterior cingulate gyrus in social valuation. *science, 313*(5791), 1310-1312.

S

Sacks, S., Cleland, C. M., Melnick, G., Flynn, P. M., Knight, K., Friedmann, P. D., . . . Coen, C. (2009). Violent Offenses Associated with Co-Occurring Substance Use and Mental Health Problems: Evidence from CJDATS. *Behavioral sciences & the law, 27*(1), 51-69. doi:10.1002/bsl.850

Salihovic, S., Kerr, M., & Stattin, H. (2014). Under the surface of adolescent psychopathic traits: high-anxious and low-anxious subgroups in a community sample of youths. *Journal of Adolescence, 37*(5), 681-689. doi:10.1016/j.adolescence.2014.03.002

Salomons, T. V., Johnstone, T., Backonja, M. M., & Davidson, R. J. (2004). Perceived controllability modulates the neural response to pain. *J Neurosci, 24*(32), 7199-7203. doi:10.1523/JNEUROSCI.1315-04.2004

Sanfey, A. G., Stallen, M., & Chang, L. J. (2014). Norms and expectations in social decision-making. Trends in cognitive sciences, 18(4), 172-174.

Santesso, D. L., & Segalowitz, S. J. (2009). The error-related negativity is related to risk taking and empathy in young men. *Psychophysiology, 46*(1), 143-152. doi:10.1111/j.1469-8986.2008.00714.x

Schmidt, M. W., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., . . . Manning, D. A. (2011). Persistence of soil organic matter as an ecosystem property. *Nature, 478*(7367), 49-56.

Schneider, A. L., & Ingram, H. M. (2005). *Deserving and entitled: Social constructions and public policy*: SUNY Press.

Schultz, D. H., Balderston, N. L., Baskin-Sommers, A. R., Larson, C. L., & Helmstetter, F. J. (2016). Psychopaths show enhanced amygdala activation during fear conditioning. *Frontiers in Psychology, 7*, 348.

Seara-Cardoso, A., Dolberg, H., Neumann, C., Roiser, J. P., & Viding, E. (2013). Empathy, morality and psychopathic traits in women. *Personality and Individual Differences, 55*(3), 328-333.

Seara-Cardoso, A., Neumann, C., Roiser, J., McCrory, E., & Viding, E. (2012). Investigating associations between empathy, morality and psychopathic personality traits in the general population. *Personality and Individual Differences, 52*(1), 67-71.

Seara-Cardoso, A., Sebastian, C. L., Viding, E., & Roiser, J. P. (2016). Affective resonance in response to others' emotional faces varies with affective ratings and psychopathic traits in amygdala and anterior insula. *Social neuroscience, 11*(2), 140-152.

Seara-Cardoso, A., Viding, E., Lickley, R. A., & Sebastian, C. L. (2015). Neural responses to others' pain vary with psychopathic traits in healthy adult males. *Cogn Affect Behav Neurosci, 15*(3), 578-588. doi:10.3758/s13415-015-0346-7

Seidel, E.-M., Habel, U., Kirschner, M., Gur, R. C., & Derntl, B. (2010). The impact of facial emotional expressions on behavioral tendencies in women and men. *Journal of Experimental Psychology: Human Perception and Performance, 36*(2), 500.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences, 1156*(1), 81-96.

Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157-1162.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*(7075), 466-469. doi:10.1038/nature04271

Skeem, J., Johansson, P., Andershed, H., Kerr, M., & Louden, J. E. (2007). Two subtypes of psychopathic violent offenders that parallel primary and secondary variants. *Journal of Abnormal Psychology, 116*(2), 395.

Skeem, J. L., & Cooke, D. J. (2010). Is criminal behavior a central component of psychopathy? Conceptual directions for resolving the debate. *Psychological Assessment, 22*(2), 433.

Skeem, J. L., Johansson, P., Andershed, H., Kerr, M., & Louden, J. E. (2007). Two subtypes of psychopathic violent offenders that parallel primary and secondary variants. *Journal of abnormal psychology, 116*(2), 395-409. doi:10.1037/0021-843X.116.2.395

Skeem, J. L., & Mulvey, E. P. (2001). Psychopathy and community violence among civil psychiatric patients: results from the MacArthur Violence Risk Assessment Study. *Journal of Consulting and Clinical psychology, 69*(3), 358.

Skeem, J. L., Poythress, N., Edens, J. F., Lilienfeld, S. O., & Cale, E. M. (2003). Psychopathic personality or personalities? Exploring potential variants of psychopathy and their implications for risk assessment. *Aggression and Violent Behavior, 8*(5), 513-546.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290-312.

Solarz, A. K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of experimental psychology, 59*(4), 239.

Steffensen, S. C., Ohran, A. J., Shipp, D. N., Hales, K., Stobbs, S. H., & Fleming, D. E. (2008). Gender-selective effects of the P300 and N400 components of the visual evoked potential. *Vision Res, 48*(7), 917-925. doi:10.1016/j.visres.2008.01.005

Stephan, K. E., Iglesias, S., Heinzle, J., & Diaconescu, A. O. (2015). Translational perspectives for computational neuroimaging. *Neuron, 87*(4), 716-732.

Stephan, K. E., & Mathys, C. (2014). Computational approaches to psychiatry. *Current opinion in neurobiology, 25*, 85-92.

Stins, J. F., Roelofs, K., Villan, J., Kooijman, K., Hagenaars, M. A., & Beek, P. J. (2011). Walk to me when I smile, step back when I'm angry: emotional faces modulate whole-body approach–avoidance behaviors. *Experimental brain research, 212*(4), 603-611.

Stoff, D. M., Breiling, J. E., & Maser, J. D. (1997). *Handbook of antisocial behavior*: John Wiley & Sons Inc.

Struijs, S. Y., Lamers, F., Vroling, M. S., Roelofs, K., Spinhoven, P., & Penninx, B. W. (2017). Approach and avoidance tendencies in depression and anxiety disorders. *Psychiatry Research, 256*, 475-481.

Sutton, R. S., & Barto, A. G. (2011). Reinforcement learning: An introduction.

T

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences, 3*(4), 151-162.

Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *Journal of Neuroscience, 33*(19), 8264-8269.

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 640-657.

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. *The SAGE handbook of personality theory and assessment, 2*, 261-292.

Terburg, D., & van Honk, J. (2013). Approach–avoidance versus dominance–submissiveness: A multilevel neural framework on how testosterone promotes social status. *Emotion Review, 5*(3), 296-302.

Tholen, D. W., Kroll, M., Astles, J. R., Caffo, A. L., Happe, T., Krouwer, J., & Lasky, F. (2003). *Evaluation of the linearity of quantitative measurement procedures: a statistical approach; approved guideline* (Vol. 23): CLSI.

Thoma, P., & Bellebaum, C. (2011). Your error's got me feeling—how empathy relates to the electrophysiological correlates of performance monitoring. *Factors mediating performance monitoring in humans–from context to personality*, 8.

Thompson, D. F., Ramos, C. L., & Willett, J. K. (2014). Psychopathy: clinical features, developmental basis and therapeutic challenges. *Journal of clinical pharmacy and therapeutics, 39*(5), 485-495.

Thompson, D. F., Ramos, C. L., & Willett, J. K. (2014). Psychopathy: clinical features, developmental basis and therapeutic challenges. *J Clin Pharm Ther, 39*(5), 485-495. doi:10.1111/jcpt.12182

Tiihonen, J., Rossi, R., Laakso, M. P., Hodgins, S., Testa, C., Perez, J., . . . Aronen, H. J. (2008). Brain anatomy of persistent violent offenders: more rather than less. *Psychiatry Research: Neuroimaging, 163*(3), 201-212.

Tillem, S., Ryan, J., Wu, J., Crowley, M. J., Mayes, L. C., & Baskin-Sommers, A. (2016). Theta phase coherence in affective picture processing reveals dysfunctional sensory integration in psychopathic offenders. *Biological psychology, 119*, 42-45.

V

van Baar, J. M., Chang, L. J., & Sanfey, A. G. (2019). The computational and neural substrates of moral strategies in social decision-making. *Nature communications, 10*(1), 1-14.

van Baar, J. M., Klaassen, F. H., Ricci, F., Chang, L. J., & Sanfey, A. G. (2020). Stationary distribution of moral strategies in a population.

Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The, 25*(4), 40.

Van Hoesen, G. W., Morecraft, R. J., & Vogt, B. A. (1993). Connections of the monkey cingulate cortex. In *Neurobiology of cingulate cortex and limbic thalamus* (pp. 249-284): Springer.

van Honk, J., & Schutter, D. J. (2006). Unmasking feigned sanity: A neurobiological model of emotion processing in primary psychopathy. *Cognitive neuropsychiatry, 11*(3), 285-306.

Venables, N. C., & Patrick, C. J. (2012). Validity of the Externalizing Spectrum Inventory in a criminal offender sample: relations with disinhibitory psychopathology, personality, and psychopathic features. *Psychological assessment, 24*(1), 88-100. doi:10.1037/a0024703

Verona, E., Patrick, C. J., & Joiner, T. E. (2001). Psychopathy, antisocial personality, and suicide risk. *J Abnorm Psychol, 110*(3), 462-470.

Vieira, J. B., Almeida, P. R., Ferreira-Santos, F., Barbosa, F., Marques-Teixeira, J., & Marsh, A. A. (2014). Distinct neural activation patterns underlie economic decisions in high and low psychopathy scorers. *Social cognitive and affective neuroscience, 9*(8), 1099-1107.

Vieira, J. B., Tavares, T. P., Marsh, A. A., & Mitchell, D. G. (2017). Emotion and personal space: Neural correlates of approach‐avoidance tendencies to different facial expressions as a function of coldhearted psychopathic traits. *Human brain mapping, 38(3)*, 1492-1506.

Villemure, C., & Bushnell, M. C. (2009). Mood influences supraspinal pain processing separately from attention. *J Neurosci, 29*(3), 705-715. doi:10.1523/JNEUROSCI.3822-08.2009

Vincent, G. M., Vitacco, M. J., Grisso, T., & Corrado, R. R. (2003). Subtypes of adolescent offenders: affective traits and antisocial behavior patterns. *Behavioral Sciences and the Law, 21*(6), 695-712. doi:10.1002/bsl.556

Vitacco, M. J., Neumann, C. S., & Jackson, R. L. (2005). Testing a four-factor model of psychopathy and its association with ethnicity, gender, intelligence, and violence. *Journal of consulting and clinical psychology, 73*(3), 466.

Volman, I., Toni, I., Verhagen, L., & Roelofs, K. (2011). Endogenous testosterone modulates prefrontal–amygdala connectivity during social emotional behavior. *Cerebral Cortex, 21*(10), 2282-2290.

Volman, I., von Borries, A. K. L., Bulten, B. H., Verkes, R. J., Toni, I., & Roelofs, K. (2016). Testosterone modulates altered prefrontal control of emotional actions in psychopathic offenders. *ENeuro, 3*(1).

Von Borries, A., Brazil, I. A., Bulten, B., Buitelaar, J., Verkes, R., & De Bruijn, E. (2010). Neural correlates of error-related learning deficits in individuals with psychopathy.

von Borries, A. K. L., Volman, I., de Bruijn, E. R. A., Bulten, B. H., Verkes, R. J., & Roelofs, K. (2012). Psychopaths lack the automatic avoidance of social threat: relation to instrumental aggression. *Psychiatry Research, 200*(2-3), 761-766.

Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K. J., & Stephan, K. E. (2014). Spatial attention, precision, and Bayesian inference: a study of saccadic response speed. *Cerebral cortex, 24*(6), 1436-1450.

Vossen, H., Van Breukelen, G., Hermens, H., Van Os, J., & Lousberg, R. (2011). More potential in statistical analyses of event‐ related potentials: a mixed regression approach. *International journal of methods in psychiatric research, 20*(3), e56-e68.

Vossen, H. G. M., van Breukelen, G. J. P., van Os, J., Hermens, H. J., & Lousberg, R. (2011). Association Between Event-Related Potentials and Pain Ratings Not as Straightforward as Often Thought. *Journal of Psychophysiology, 25*(1), 18-25. doi:10.1027/0269-8803/a000028

W

Wallace, J. F., Malterer, M. B., & Newman, J. P. (2009). Mapping Gray's BIS and BAS constructs onto factor 1 and factor 2 of Hare's psychopathy checklist–revised. *Personality and Individual Differences, 47*(8), 812-816.

Waller, R., Gard, A. M., Shaw, D. S., Forbes, E. E., Neumann, C. S., & Hyde, L. W. (2019). Weakened functional connectivity between the amygdala and the ventromedial prefrontal cortex is longitudinally related to psychopathic traits in low-income males during early adulthood. *Clinical Psychological Science, 7(3)*, 628-635.

Walton, M. E., Devlin, J. T., & Rushworth, M. F. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nature neuroscience, 7*(11), 1259-1265.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson, 22*, 498.

Weinstein, E., Feldman, K. A., Goodman, N., & Markowitz, M. (1972). Empathy and communication efficiency. *The Journal of social psychology, 88*(2), 247-254.

Welker, K. M., Lozoya, E., Campbell, J. A., Neumann, C. S., & Carré, J. M. (2014). Testosterone, cortisol, and psychopathic traits in men and women. *Physiology & behavior, 129*, 230-236.

Weller, J. A., Levin, I. P., Shiv, B., & Bechara, A. (2009). The effects of insula damage on decision-making for risky gains and losses. *Social neuroscience, 4*(4), 347-358.

Welsh, G. S. (1952). An anxiety index and an internalization ratio for the MMPI. *Journal of Consulting & Clinical Psychology, 16*(1), 65-72.

Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS quarterly*, 177-195.

Widiger, T. A., & Sankis, L. M. (2000). Adult psychopathology: issues and controversies. *Annual Review of Psychology, 51*, 377-404. doi:10.1146/annurev.psych.51.1.377

Williams, K. M., Paulhus, D. L., & Hare, R. D. (2007). Capturing the four-factor structure of psychopathy in college students via self-report. *Journal of personality assessment, 88*(2), 205-219.

Williamson, S., Harpur, T. J., & Hare, R. D. (1991). Abnormal processing of affective words by psychopaths. *Psychophysiology, 28*(3), 260-273.

Woltering, S., Bazargani, N., & Liu, Z.-X. (2013). Eye blink correction: A test on the preservation of common ERP components using a regression based technique. *PeerJ, 1*, e76.

Wong, S. C.-P., & Hare, R. D. (2006). *Guidelines for a psychopathy treatment program.*

Wootton, J. M., Frick, P. J., Shelton, K. K., & Silverthorn, P. (1997). Ineffective parenting and childhood conduct problems: the moderating role of callous-unemotional traits. *Journal of consulting and clinical psychology, 65*(2), 301.

Y

Yang, Y., Raine, A., Colletti, P., Toga, A. W., & Narr, K. L. (2010). Morphological alterations in the prefrontal cortex and the amygdala in unsuccessful psychopaths. *Journal of Abnormal Psychology, 119*(3), 546.

Yau, W.-Y. W., Zubieta, J.-K., Weiland, B. J., Samudra, P. G., Zucker, R. A., & Heitzeg, M. M. (2012). Nucleus accumbens response to incentive stimuli anticipation in children of alcoholics: relationships with precursive behavioral risk and lifetime alcohol use. *Journal of Neuroscience, 32*(7), 2544-2551.

Yildirim, B. O., & Derksen, J. J. (2012). A review on the relationship between testosterone and the interpersonal/affective facet of psychopathy. Psychiatry Research, 197(3), 181-198.

Yoshida, W., & Ishii, S. (2006). Resolution of uncertainty in prefrontal cortex. *Neuron, 50*(5), 781-789.

Z

Zhao, K., & Smillie, L. D. (2015). The role of interpersonal traits in social decision making: Exploring sources of behavioral heterogeneity in economic games. *Personality and Social Psychology Review, 19*(3), 277-302.

Nederlandse Samenvatting

Psychopathische persoonlijkheidstrekken en mechanismen onderliggend aan antisociaal keuzegedrag

De meeste mensen zijn bekend met de term ‘psychopaat’, welke in de volksmond vaak gebruikt wordt om iemand te beschrijven die moordlustig, koelbloedig, en moreel gestoord is. Deze beeldvorming van psychopathie en de populariteit ervan is terug te zien in een grote verscheidenheid aan romans, films, tv-series en computerspellen waarin een beruchte psychopaat de hoofdrol speelt, maar ook in de manier waarop de populaire media de term gebruikt om te refereren naar moordlustige criminelen. Hierdoor is in de maatschappij het beeld ontstaan dat psychopathie vrijwel altijd gepaard gaat met deze extreme gedragsuitingen. De definitie van psychopathie zoals gangbaar in de psychiatrie, criminologie en in wetenschappelijk onderzoek is echter genuanceerder. Psychopathie wordt hierin gedefinieerd als een persoonlijkheidsstoornis die gepaard gaat met verstoringen op sociaal-emotioneel vlak in combinatie met antisociaal gedrag. Bovendien blijkt uit onderzoeksgegevens dat psychopathie niet uitsluitend voorkomt in penitentiaire of forensisch-psychiatrische populaties, maar dat psychopathische kenmerken ook aanwezig zijn in de algemene populatie (bij ongeveer 1 tot 6%). Hoewel de uitingen van psychopathie over het algemeen minder extreem zijn in de algemene populatie, komen de kenmerken kwalitatief overeen met de psychopathische kenmerken in penitentiaire of forensisch-psychiatrische populaties.

Deze definitie van psychopathie komt van de Canadese psycholoog en hoogleraar Robert Hare. Hij beschrijft het construct aan de hand van vier kenmerken, te weten afwijkingen op interpersoonlijk, affectief, en antisociaal vlak in combinatie met een grillige levensstijl. Het meetinstrument dat aansluit op zijn vierfactorenmodel is de *Psychopathic Checklist – Revised* (PCL-R). De PCL-R omvat een semi-gestructureerd interview met de betreffende persoon en een analyse van overige relevante informatie. Afname vereist klinische training en toegang tot andere relevante informatie en is daarom niet geschikt om psychopathie in de algemene populatie te onderzoeken. Er zijn verschillende (zelf)rapportage-instrumenten ontwikkeld die wel gebruikt kunnen worden om psychopathische persoonlijkheidskenmerken te meten in de algemene populatie, waarvan de *Self-Report Psychopathy Scale* (SRP) de meestgebruikte is. Het voordeel van dit instrument is dat deze dezelfde structuur aanhoudt als de PCL-R en ook uitgaat van het vierfactorenmodel van Hare.

Binnen de klinische psychiatrie en het wetenschappelijk onderzoek wordt vaak onderscheid gemaakt tussen verschillende soorten psychopathie, waarbij een onderscheid tussen primaire en secundaire subtypes gemaakt wordt. Primaire psychopathie wordt specifiek gekenmerkt door emotieloos, berekenend en manipulatief gedrag, instrumentele agressie en minder angstgevoelens. Secundaire psychopathie wordt gekenmerkt door impulsiviteit, reactieve agressie en angst. In hoofdstuk 2 van dit proefschrift is onderzocht of verschillende subtypes herkend konden worden bij een gevangenispopulatie in Wisconsin in de Verenigde Staten, op basis van resultaten van de Self-Report Psychopathy Scale. Door gebruik te maken van een specifieke statistische methode (zogenaamde latente-profiel-analyse) bleek het mogelijk om vier verschillende subtypes kunnen onderscheiden. De grootste groep, die we *generic offenders* (algemene delinquenten) noemden, bestond uit gedetineerden die relatief laag scoorden op alle vier de factoren van de SRP. De tweede groep, welke we de impulsief-antisociale groep (*impulsive-antisocial*) noemden, scoorde relatief hoog op de factoren 'leefstijl' en 'antisociaal gedrag', terwijl de scores op de andere twee factoren relatief laag waren. De derde groep, niet-antisociale delinquenten met psychopathische kenmerken (*non-antisocial psychopathic traits offenders*), scoorde relatief laag op de factor 'antisociaal gedrag', terwijl de scores op de andere factoren relatief hoog waren. De groep die hoog scoorde op alle vier de factoren noemden we delinquenten met psychopathische trekken (*psychopathic traits offenders*). We hebben de vier subtypes verder kunnen specificeren aan de hand van vragenlijsten die andere persoonlijkheidskenmerken en gedrag dat vaak in verband wordt gebracht met antisociale populaties meten. Deze bevindingen komen overeen met resultaten uit eerdere studies die subtypes onderscheidden op basis van PCL-R-scores. Mijn onderzoek ondersteunt derhalve het onderscheid in subtypes binnen een gevangenispopulatie. Bovendien laat deze studie zien dat de SRP bruikbaar en valide is om antisociale subtypes vast te stellen bij een gevangenispopulatie.

Recente technologische ontwikkelingen hebben ervoor gezorgd dat we in een relatief korte tijd meer inzicht hebben gekregen in de onderliggende cognitieve en neurobiologische mechanismen van verschillende persoonlijkheidsstoornissen, waaronder psychopathie. Dit heeft geleid tot verschillende perspectieven die psychopathie niet alleen willen verklaren op basis van persoonlijkheidskenmerken, maar ook op grond van onderliggende cognitieve en neurobiologische factoren. Volgens een veelgebruikt neurocognitief model, het zogenaamde *Integrated Emotions System* (IES-model), vormen verstoringen op sociaal-emotioneel vlak en afwijkend keuzegedrag de kerneigenschappen van psychopathie. Deze kerneigenschappen worden in verband gebracht met specifieke afwijkingen in de hersenen, bijvoorbeeld in de amygdala, de ventromediale prefrontaalkwab en het striatum.

De overige hoofdstukken van dit proefschrift kunnen in twee delen worden verdeeld. Het eerste deel omvat twee hoofdstukken waarin sociaal-affectief functioneren in relatie tot psychopathische kenmerken onderzocht is. Het IES-model suggereert dat psychopathisch gedrag veroorzaakt wordt door verstoringen in specifieke affectief-associatieve leerprocessen in de amygdala. Onderzoek binnen gevangenispopulaties lieten een link zien tussen psychopathie en verstoringen in de verwerking van signalen van dreiging. In hoofdstuk 3 is onderzocht hoe volwassen proefpersonen met uiteenlopende scores op de SRP reageerden op confrontatie met blije en boze gezichtsuitdrukkingen. Verder wilden we weten of dit gedrag samenhing met het testosteronniveau van de proefpersonen. Over het algemeen zijn mensen geneigd om toenadering te zoeken bij confrontatie met iets positiefs (bijvoorbeeld een blij gezicht) en afstand te nemen bij confrontatie met iets negatiefs (bijvoorbeeld een boos gezicht). In een experimentele setting is dit gedrag gemeten door te kijken naar de snelheid waarmee mensen een joystick naar zich toe of van zich af bewegen wanneer ze en blij of boos gezicht op een computerscherm zien. De resultaten uit deze studie toonden aan dat mensen die hoog scoren op kenmerken van psychopathie minder geneigd zijn om afstand te nemen wanneer ze geconfronteerd worden met dreiging en zelfs meer geneigd zijn om toenadering te zoeken. Het testosteronniveau van de proefpersonen bleek echter geen verband te hebben met dit gedrag.

Een andere affectieve modaliteit die vaak in verband wordt gebracht met psychopathie, is de verwerking van pijnprikkels. Hoofdstuk 4 beschrijft een studie waarin de pijngevoeligheid en de empathie onderzocht zijn door gebruik te maken van metingen van specifieke hersengolven. Bovendien is onderzocht of dit gedrag samenhing met de score op de SRP. Proefpersonen werden geïnstrueerd om achtereenvolgens de rol van 'schurk' aan te nemen, waarbij ze een andere persoon observeerden terwijl deze elektrische schokken ontving, en 'slachtoffer', waarbij ze zelf schokken ontvingen terwijl een andere persoon toekeek. Daarnaast werd de controle over de knop waarmee een schok werd gegeven gemanipuleerd, waarbij 'passief' verwijst naar het hebben van geen controle over de schokken, terwijl 'actief' verwijst naar controle hebben over het toedienen van de schokken. De resultaten lieten zien dat individuen meer conflicten ervoeren wanneer ze iemand anders pijn deden dan wanneer ze zichzelf pijn deden. In tegenstelling tot onze verwachting bleek dit effect onafhankelijk te zijn van de mate van psychopathische kenmerken. Verder gaven de resultaten aan dat zelfgecontroleerde pijn als pijnlijker werd ervaren dan ongecontroleerde pijn. Dit effect bleek negatief gerelateerd te zijn aan de mate van psychopathische eigenschappen. Deze relatie zou verklaard kunnen worden door de verminderde gevoeligheid voor pijn die in eerdere onderzoeken in verband is gebracht met psychopathie.

Naast de affectieve stoornissen, stelt het IES-model dat afwijkend keuzegedrag een centraal kenmerk van psychopathie is. De huidige samenleving heeft complexe sociale structuren; in de meeste gevallen hebben de keuzes die we maken dan ook invloed op mensen in onze omgeving. Psychopathische eigenschappen worden in verband gebracht met antisociaal keuzegedrag; mensen die hoog scoren op psychopathische kenmerken hebben vaak geen oog voor de impact die hun beslissingen hebben op het welzijn van anderen. In het tweede deel van dit proefschrift zijn twee studies beschreven die de mechanismen van besluitvorming en het verband met psychopathische eigenschappen onderzochten. Hoofdstuk 5 beschrijft een studie waarin is onderzocht hoe mensen gebruik maken en leren van sociale en niet-sociale informatie. Deelnemers deden een taak waarin ze steeds moesten kiezen tussen een blauwe en een groene kaart, waarbij een van de kaarten correct was. Ze konden gebruik maken van twee bronnen van informatie, 1) de uitkomst van voorgaande trials (niet-sociale informatie of wel *reward* informatie) en 2) advies wat werd gegeven door een andere speler (sociale informatie). De betrouwbaarheid van deze twee bronnen van informatie wisselde zich af, soms in een snel tempo en soms in een langzaam tempo. De afweging tussen deze twee bronnen van informatie was van invloed is op de beloning die deelnemers konden verdienen. De mate waarin deelnemers risico namen werd bijgehouden door deelnemers iedere ronde opnieuw zelf hun inzet te laten bepalen. Met behulp van een computationeel model kon worden bepaald hoe psychopathische persoonlijkheidstrekken gerelateerd waren aan het gebruik van sociale en niet-sociale informatie. Ook zijn hersengolven met een specifieke frequentie (van 4 tot 8Hz, zogenaamde thetagolven) gemeten, omdat uit eerder onderzoek is gebleken dat theta-activiteit samenhing met adaptieve controleprocessen in de hersenen. De bevindingen van hoofdstuk 5 suggereerden dat, hoewel het leren van de niet-sociale informatie intact was, het leren van sociale informatie negatief geassocieerd was met psychopathische eigenschappen. Mensen met veel interpersoonlijke eigenschappen konden zich minder goed aanpassen aan veranderingen in de betrouwbaarheid van sociale informatie. Hoewel dit een negatieve invloed had op hun prestaties op de taak, veranderde het niet de mate waarin risico werd genomen of de mate waarin deelnemers voorkeur hadden voor een van de bronnen van informatie. De theta-hersengolven waren negatief gecorreleerd aan psychopathische kenmerken, wat aansluit bij eerdere bevindingen die lieten zien dat theta-activiteit een rol speelt bij het inschatten van de betrouwbaarheid van sociale informatie. De bevindingen van deze studie benadrukken de relatie tussen verminderd sociaal-affectief functioneren en het effect ervan op besluitvorming in een sociale context.

Hoofdstuk 6 beschrijft een experiment waarin morele strategieën (dat wil zeggen beslissingsstijlen) werden bestudeerd in een taak waarbij keuzegedrag van de deelnemer gevolgen heeft voor zowel de deelnemer zelf als de (onbekende) bondgenoot. Deelnemers ontvingen iedere ronde een aantal munten, waarbij de hoeveelheid afhankelijk was van de inzet van de bondgenoot en een vermenigvuldigingsfactor. De taak voor de deelnemers was om de munten te verdelen over zichzelf en de bondgenoot. Deelnemers hadden informatie over de vermenigvuldigingfactor, maar ze wisten ook dat de bondgenoot er vanuit gingen dat de vermenigvuldigingsfactor altijd gelijk was. Een computationeel model werd gebruikt om de rol van verschillende strategieën in deze taak te schatten en onderzocht hoe dit verband hield met psychopathische eigenschappen. Deze resultaten gaven aan dat er bij deze taak vijf verschillende morele strategieën werden gebruikt; *guilt aversion* (aversie voor schuldgevoel), *inequity aversion* (aversie voor ongelijkheid), *moral opportunism* (moreel opportunisme), *greed* (hebzucht) en *generosity (*vrijgevigheid). Twee parameters die *guilt aversion* en *inequity aversion* vertegenwoordigen, werden gemeten in relatie tot het niveau van psychopathische eigenschappen van de deelnemers. *Inequity aversion* in deze taak weerspiegelde de neiging om een gelijkmatige verdeling tussen de deelnemer en de bondgenoot te verzekeren. *Guilt aversion* weerspiegelde de neiging om de verwachtingen van anderen te volgen om schuldgevoelens te vermijden. De resultaten lieten zien dat *inequity aversion*, maar niet *guilt aversion*, samenhing was met lagere scores op de affectieve en antisociale factoren van de SRP. Bovendien gaven de resultaten aan dat een hogere mate van antisociale eigenschappen samenhing met een lagere hoeveelheid munten dat aan de bondgenoot werd teruggegeven. Personen met veel antisociale trekken handelen dus mogelijk hebzuchtiger in een sociaaleconomisch spel, wat aansluit bij eerdere bevindingen in de literatuur. Op basis van deze resultaten kan geconcludeerd worden dat een verminderd gevoel voor eerlijkheid mogelijk de onderliggende motivatie is voor hebzuchtig gedrag bij mensen die hoog scoren op affectieve en antisociale kenmerken.

Het onderzoek dat in dit proefschrift wordt gepresenteerd biedt nieuwe inzichten in mogelijke mechanismen die ten grondslag liggen aan de verstoringen op sociaal-emotioneel vlak en op het gebied van besluitvorming die gerelateerd zijn aan kenmerken van psychopathie. Bovendien benadrukken de resultaten het belang van de noodzaak tot het verbeteren van onderzoekstechnieken. Hierdoor kunnen neurocognitieve processen die ten grondslag liggen aan psychopathisch gedrag beter onderzocht worden, waardoor bevindingen en perspectieven uit verschillende onderzoeksgebieden geïntegreerd kunnen worden.

Dankwoord

Er zijn een heleboel mensen die mij direct dan wel indirect hebben gesteund en geholpen bij het uitvoeren van mijn onderzoek en het schrijven van dit proefschrift. Ik wil hen bij deze hiervoor bedanken, zonder jullie was het me niet gelukt om dit voor elkaar te krijgen! Een aantal van hen wil ik in het bijzonder bedanken.

First of all, I would like to thank my promotors Roy and Jan and co-promotors Inti and Jeffrey for all the effort they have made to make it possible for me to carry out my PhD project. Roy en Jan, ik wil jullie ontzettend bedanken voor jullie vertrouwen, bereidheid om mee te denken en dat ik bij jullie terecht kon op de momenten wanneer dat nodig was. Jeffrey, I'm still happy today that I've contacted you for an internship after you inspired me with your lectures about psychopathy and antisocial behavior during one of the CNS master courses. After the master, you've created the opportunity for me to start a PhD on these topics. During my PhD, there were moments in which you were present and moments in which you were more at the background; I'm thankful for the opportunities you gave me, for sharing your insights, and for having faith in me. En dan Inti; zonder jou was ik niet geweest waar ik nu ben en het is daarom een uitdaging om je in een paar woorden te bedanken. Ik heb ontzettend veel van je geleerd, ik kon altijd op je rekenen, je wist me op de goede momenten een duwtje in de rug te geven, maar ook wanneer je me beter even kon laten. Bedankt voor het vertrouwen dat je me hebt gegeven, je vrolijkheid en je talent om me te overtuigen dat het allemaal wel goed komt. Een betere begeleider had ik me niet kunnen wensen!

Furthermore, I would like to thank all the other co-authors for their indispensable contribution to the manuscripts that have emerged from my research. In het bijzonder; Marijtje, Tineke en Casper, bedankt dat jullie me mee hebben genomen in het uitdagende en leerzame Discovery project; co-members of the SSSP community, thank you so much for sharing your time and expertise; Jeroen, bedankt voor je enthousiasme en behulpzaamheid, ik neem me voor om gebruiks-vriendelijke scripts te schrijven, jij doet het ook echt! Andreea, visiting you and your lab in Basel was inspiring, thank you for being so kind and helpful, and thank you for your patience when teaching me to use your complicated models.

I would like to thank my interns for their enthusiasm, their insights and their help in data collection. In particular, Ira; you are one of a kind, I will never forget your entrances in the lab on our days of testing, always in a rush, carrying a ton of stuff with you, but ever polite and motivated to get a lot of work done.

Verder wil ik alle (oud) NRP-collega's bedanken voor de fijne tijd die ik met hen heb gehad. In het bijzonder Nikki; jij bent natuurlijk meer dan een collega, heerlijk hoe onze gesprekken over onderzoek meestal uitliepen in gesprekken over Trekvogels, fantasy-opstellingen, of knutselperikelen. Selma, ik kan me vooral mijn voeten in Krakau en jouw voeten tijdens de Kennedymars herinneren. Saskia, fijn hoe je altijd bereid bent om te helpen en mee te denken, bedankt ook voor je steun. Dear SAMBA dancers, it's awesome to see how the lab grew into an actual group over the last years and I'm proud that I could be part of it! Johanna, ik vond het altijd gezellig als je plotseling weer eens binnen kwam wandelen en we even konden klagen over van alles en nog wat. Ivana and Esther, I enjoyed the time you both were here and we went for good food and beers. Thanks to my fellow PhD students of the Translational Psychiatry lab for the good times and Shaha and Joanna for teaching me some of their labskills.

Syanah, dit is niks voor ons, maar ik ga hier toch van de gelegenheid gebruik maken om een aantal dingen te noemen die mijn PhD tijd extra kleur hebben gegeven, en daar heb jij een grote rol in gespeeld. Sportlesjes, gebarenchallenge, paaldanslessen, ImpStim, push-up records, pannenkoeken, Neuronusreisjes, Glasgowtrip, matlabsessies, Kennedymars, Bruggenloop, BKO+wijn, SPAR-wandelingen, RustyLake, en nog veel meer; we hebben zeker 5 jaar een kamer gedeeld en bleken meer gemeen te hebben dan we in eerste instantie dachten. Bedankt voor je bijdrage aan mijn onderzoek en voor alle gezelligheid, we keep in touch! Xiaochen, although we did the same master (or not?;)), our friendship grew over the last years of our PhDs. I enjoyed your sudden knocks at our office door, your colorful appearance, (random) stories, criticism and complaints, Chinese treats, riddles, and birthday cakes. I hope the future bring us more sport try-outs and above all, sushi!

Lieve Taartjes, onze vriendschap is ontstaan in de snijzaal tijdens de fascinerende anatomielessen. Het is mooi om te zien hoe iedereen daarna zijn weg heeft gevonden en ik ben blij dat we onze avonturen nog steeds delen!

Grote dank gaat uit naar alle proefpersonen die deel hebben genomen aan mijn onderzoek, zonder hen was dit alles uiteraard niet mogelijk geweest.

Naast iedereen die een belangrijke rol heeft gespeeld in mijn promotietraject, zijn er ook een hele hoop vrienden en familie die ik wil bedanken voor hun interesse, luisterend oor, maar vooral ook voor alle leuke en minder leuke dingen in het leven die ik met hen kan delen.

Lieve Groenteburgers, we kennen elkaar al heeeel lang, onze urenlange lunches zijn altijd iets om naar uit te kijken! Teamies, Trekvogels en verwante vrienden en vriendinnen, voetbal en vriendschap gaan heel goed samen, dat is gebleken! Strijders, we delen herinneringen aan memorabele avonturen! An-mei, altijd oprechte interesse, proost op nog vele serieuze en minder serieuze gesprekken onder het genot van een biertje! En dan Lola en Charlie, jullie komen in alle groepen terug en dat is niet voor niks, dank voor onze vriendschap en dat jullie er altijd voor mij zijn!

Lieve familie en schoonfamilie, heel erg bedankt voor jullie liefde en gezelligheid, jullie interesse in werk en welzijn, en jullie steun in moeilijke tijden. In het bijzonder, Nura; we zagen elkaar natuurlijk sowieso al veel, maar het was extra gezellig dat we ook de lunch en koffiepauzes op de RU konden delen. Elsa; wat een pech, nog geen promotiefeest, misschien komt het ooit nog. Loesje; ervaringsdeskundige, fijn dat jij onzekerheden en frustraties rondom een PhD altijd zo goed kon begrijpen. En dan Roos en Nath; bij jullie kan ik altijd terecht voor gezelligheid (en een lekker hapje), we hebben een hoop meegemaakt met z'n vieren en het is fijn om te weten dat we er altijd voor elkaar zijn, jullie zijn onmisbaar! Papa; ik had graag je reactie gehoord na afloop van m'n promotie, ik denk dat Rosalie en ik daar flink om zouden kunnen lachen. Ik ben blij dat ik, ook tijdens een groot deel van mijn promotietraject, dicht bij je heb kunnen zijn. Mama; vanuit jouw werk wist jij heel goed wat er allemaal bij een PhD komt kijken, maar je kende me gelukkig ook zo goed dat je wist dat je je er niet altijd mee moest bemoeien. Die drang naar zelfstandigheid en nuchterheid heb ik van jou, dat ben ik de afgelopen tijd des te meer gaan beseffen. Jouw liefde en trots zijn niet aan mij voorbij gegaan, die houd ik voor altijd bij me.

En tot slot lieve Koen; jij hebt mijn promotietraject van dichtbij meegemaakt en je hebt daarin meer betekend dan dat je waarschijnlijk zelf weet. Jouw energie, positiviteit en enthousiasme zijn aanstekelijk en werken relativerend, bedankt voor je liefde, je begrip en dat je er altijd voor me bent, we gaan samen nog heel veel mooie dingen beleven!

Curriculum Vitae

Josi Driessen was born on the 8th of December, 1989 in Nijmegen. She graduated from the Montessori College in Nijmegen and continued with the bachelor Psychobiology at the University of Amsterdam. During the third year of this bachelor, she did her first research internship at the Radboud University in Nijmegen on comorbidity of absence epilepsy and autistic-like behavior in rats. After obtaining her bachelors degree, she continued with the research master Cognitive Neuroscience at the Radboud University in Nijmegen. She wrote her masterthesis on validity and applicability of the Self-Report Psychopathy checklist in a population sample. Following this internship, she worked as a research assistant on the Food and Cognition Model project in which she performed EEG analyses for a study on the effect of tyrosine on reponse inhibition in healthy older adults. In 2016 she started her PhD at the Donders Institute for Brain, Cognition and Behavior (Radboudumc/Radboud University). During her PhD she studied some of the mechanisms that are believed to play a role in the cognitive disruptions seen in individuals with high levels of psychopathic traits. She combined behavioral, electrophysiological and computational approaches to study affective and social behavior in relation to psychopathy in a series of experiments. Beyond her main studies, she had the opportunity to collaborate on projects using non-invasive brain stimulation and epigenetics, assist in testing in a correctional facility, and obtain her University Teaching Qualification at the Radboud University. In March 2020, Josi joined the Goal lab at Utrecht University as a postdoctoral researcher. Her current work investigates the relationship between personal autonomy and sense of agency in offenders and healthy adults.

List of publications

Driessen, J. M. A., van Baar, J. M., Sanfey, A. G., Glennon, J. C., & Brazil, I. A. (in press). Moral strategies and psychopathic traits. *Journal of Abnormal Psychology.*.

Driessen, J. M. A., Brazil, I. A., Dorta Lorenzo, E., Herwaarden, A. E., Olthaar, A. J., Potamianou, H., & Glennon, J. C. (in press). Psychopathic traits influence threat avoidance in a community sample independent of testosterone. *Personality Disorders: Theory, Research, and Treatment.*

Wynn, S. C.*, Driessen, J. M. A.*, Glennon, J. C., Brazil, I. A., Schutter, D. J. L. G. (2019). Cerebellar tDCS improves reactive response inhibition in healthy volunteers. *The Cerebellum, 18,* 983–988. **shared first-authorship*

Driessen, J. M. A., Fanti, K. A., Glennon, J. C., Neumann, C. S., Baskin-Sommers, A. R., Brazil, I. A. (2018). A comparison of latent profiles in antisocial male offenders. *Journal of Criminal Justice, 57,* 47-55.

van Heck, C. H.*, Driessen, J. M. A.*, Amato, C. H., van den Berg, M., Bhandari, P., … Jongsma, M. L. A. (2017). Pain processing in a social context and the link with psychopathic personality traits – an event-related potential study. *Frontiers in Behavioral Neuroscience*, 11, 180. **shared first-authorship*

Molleman, P.W., Driessen, J.M.A., Schilder, C., Bulten, B.H., & Brazil, I. A. (revision submitted). The predictive value of executive functioning and motivational systems in criminal behavior.

Buric, I., Farias, M., Driessen, J. M. A., Brazil, I. (submitted). Variation in meditation responses: A meta-analysis.

Smeijers, D., Uzieblo, L., Glennon, J. C., Driessen, J. M. A., Brazil, I. A. (submitted). Examining Individual Differences in Social Reward Valuation: A Person-based Approach.

Driessen, J. M. A., Diaconescu, A., Buitelaar, J. K., Kessels, R. P. C., Glennon, J. C., & Brazil, I. A. (in preparation). Dissecting how psychopathic traits are linked to disturbed learning from reward and social advice.

Gunschera, L., Brazil, I. A., Driessen, J. M. A. (in prep). Socio-economic decision-making and psychopathic traits: meta-analysis and systematic review.

Research Data Management

This research followed the applicable laws and ethical guidelines. Research Data Management was conducted according to the FAIR principles. The paragraphs below specify in detail how this was achieved.

Ethics

This thesis is based on the results of human studies, which were conducted in accordance with the principles of the Declaration of Helsinki. All participants were tested with the approval of the local ethics committee (ECSW2017-0805-512, amendment ECG2012-1301-010a2) or the Ethics Committee of the University of Wisconsin (IRB SE-2011-0358).

Funding

The work presented in this thesis has been supported by (i) the European Union Innovative Medicine Initiative (IMI) grant PRISM (n° 115916), (ii) the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 603016 (MATRICS), and (iii) the Department of Cognitive Neuroscience of Radboudumc.

Findable, Accesible

Data of chapters 4, 5, and 6 is stored at the Donders Repository and remains available for at least 10 years after termination of the studies. Informed consent was obtained on paper following the procedures of the Donders Centre of Cognition. The forms are archived in the central archive for 15 years after termination of the studies.

Interoperable, Reusable

The raw data and the analysis scripts of chapters 4, 5, and 6 are stored at the Donders Repository (https://doi.org/10.34973/db8v-s911). The raw data and analysis scripts of chapters 2 and 3 can be provided upon request (jmadriessen@gmail.com). A description of the experimental setups can be found in the published articles or in the chapters of this thesis.

Privacy

The privacy of the participants in this thesis has been warranted using individual subject codes. A pseudonymization key links this code to personal data. This key was stored on a network drive that was accessible to members of the project who needed access to it because of their role within the project. The key was stored separately from the research data.

Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology, specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy, higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses, please visit: http://www.ru.nl/donders/graduate-school/phd/

DONDERS INSTITUTE

Radboud University

Radboudumc