Impairments in social interaction of individuals with personality disorders and dark personality traits (Beperkingen in sociale interactie van individuen met persoonlijkheidsstoornissen en donkere persoonlijkheidskenmerken)

Impairments in Social Interaction of Individuals with Personality Disorders and Dark Personality Traits

Haang Jeung-Maarse

The work presented in this thesis was performed at the Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, the Netherlands, and at the Faculty of Economics and Social Sciences, Heidelberg University, Heidelberg.

ISBN: 978-94-6469-536-6

Cover illustration: Stained-glass window in the style of the Early Northern Renaissance ("Limburgse school") of three women engaging in an economic-exchange game.

Layout and printing: proefschriftmaken.nl

IMPAIRMENTS IN SOCIAL INTERACTION OF INDIVIDUALS WITH PERSONALITY DISORDERS AND DARK PERSONALITY TRAITS

Beperkingen in sociale interactie van individuen met persoonlijkheidsstoornissen en donkere persoonlijkheidskenmerken

DISSERTATION

to obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus,
Prof.dr. Pamela Habibović
in accordance with the decision of the Board of Deans,
to be defended in public
on Monday, 18 September 2023, at 10:00 hours

by

Haang Jeung-Maarse

Supervisors:
Prof.dr. K. Schruers
Prof.dr. C. Schwieren, Heidelberg University, Germany

Assessment Committee:
Prof.dr. B.P.F. Rutten (chair)
Prof.dr. J.M.L.G. à Campo
Prof.dr. S. Claes, KU Leuven, Belgium
Prof.dr. B. Dandachi-Fitzgerald
Prof.dr. M. Dudeck, University Hospital Ulm, Germany

Financial Support:
European Union's Seventh Framework Programme (FP7), grant agreement number 602407 (FemNAT-CD)
Foundation of the federal state bank of Baden-Wuerttemberg (Landesbank Baden-Wuerttemberg, LBBW), grant agreement number 2019050193
Medical Faculty of Heidelberg

"Dit is het enige wat telt, lieverd, dat iemand meer in je ziet
dan je wist dat er te zien was"

Arthur Japin, 2003

TABLE OF CONTENTS

CHAPTER 1

General introduction

Parts of the general introduction have been extracted and adapted from reviews by Jeung-Maarse & Herpertz, 2020, Herpertz, Bertsch & Jeung, 2017, Jeung et al., 2016, and Jeung et al., 2014.

PERSONALITY DISORDERS

Personality disorders (PDs) are characterized by impairments in personality functioning affecting the self-functioning level and the interpersonal functioning level. The impairments in personality functioning are central to PDs distinguishing them from healthy personality as well as from other forms of mental disorders[1]. The impairments must exist for a period of at least two years. The patterns of experience and behavior are so disturbed that personal performance in everyday social, professional and private life is difficult for individuals suffering from PDs. In the next step of the diagnostic procedure, the severity of the disorder is scaled in three stages based on the extent of the functional impairment and its clinical manifestations (cognitive, emotional and behavioral) as well as the degree of penetration in different areas of life: The dimensional degree of severity leads to the classification into the three categories of severity (mild, moderate and severe PD).

In addition to the degree of severity, the individual differences between personality disorders ("species") are then further defined by pathological personality traits. Derived from the five-factor model of personality[2], the following characteristics are distinguished: negative affectivity, dissociality, disinhibition, anankastia and detachment. "Negative affectivity" means the tendency to react frequently and inappropriately with negative emotions and includes problems in emotion regulation. Individuals who score high on negative affect are negativistic, distrustful, and exhibit low self-esteem. "Dissociality" includes selfishness and a lack of empathy. "Disinhibition" means impulsive behavior, high distractibility, irresponsibility and lack of planning. "Anankastia" includes perfectionist behavior, rigidity, and hedging behavior in decision-making situations. "Detachment" subsumes social and emotional detachment, especially avoidance of relationships, aloofness, and lack of emotional experiences.

Further, a "borderline pattern descriptor" can be specified, which is characterized by instability in interpersonal relationships, in self-image and emotions, as well as by impulsiveness and self-damaging behavior. Highly consistent with prior classification systems, borderline personality disorder (BPD) is characterized by desperate efforts to avoid actual or experienced abandonment; relationships are usually intense but unstable, as is self-image, and mood highly reactive with inadequately intense feelings of anger and also emptiness. Transient dissociative symptoms and psychosis-like symptoms occur in states of high tension.

PDs are not uncommon: Almost 10% of the total population suffer from PD[3]. While a point prevalence of 1% is estimated for BPD, 22% of patients in psychiatric clinics and 12% of outpatients suffer from BPD[4]. Results from long-term studies show that after a longer course of the disease, only a few patients still meet the criteria for PD. The remission rate for the entire PD group, depending on the

specific personality traits, is between 40 and 60% within two years[5]. As many as 90% of patients show symptomatic remission within 10 years, with only 50% having both symptomatic remission and psychosocial recovery[6]. Only 80% of patients with BPD meet criteria of good functioning[7], and only a third are even able to work full-time. Nevertheless, personal performance in everyday life remains significantly restricted, which is why social integration outside of the psychiatric care system should also be considered in the long term. The presence of a PD turns out to be a stronger predictor of negative quality of life than sociodemographic variables, somatic and other mental illnesses[8].

CLASSIFICATION SYSTEMS

This definition of Personality Disorders (PDs) refers to the latest version of the International Statistical Classification of Diseases and Related Health problems (ICD) which is the global standard for encoding medical diagnoses. Originating in the 19th century, the fundamental 11th revision of the ICD, ICD-11, was adopted by the 72nd World Health Assembly in 2019 and came into effect on 1st January 2022. Since then, World Health Organization (WHO) member states have been able to report their mortality data to the WHO using ICD-11 coding. However, there is a flexible transition period of at least 5 years. For instance, the specific date for introducing the ICD-11 in Germany and in the Netherlands for mortality coding has not yet been determined.

The ICD-11 includes a fundamental change in the diagnostic procedure, since a dimensional severity assessment takes the place of specific personality disorder categories[9,10]. This means that, apart from borderline personality disorder (BPD), no other personality disorder (PD) is specified in more detail. The previous categorical classification according to ICD-10 and diagnostic and statistical manual of mental disorders 4th/5th revision (DSM-IV/DSM-5) of the American Psychiatric Association (APA) includes, among other things, arbitrary diagnostic thresholds[11], the inflationary use of personality disorder diagnoses with an average of three comorbid diagnoses and the frequent use of an unspecific diagnostic category (so-called "personality disorder, not otherwise specified"), which is of little help for therapy planning[12], and criticized for insufficient evidence for the specific personality disorders themselves[13]. Exceptions are the studies on psychopathy (with reference to the antisocial/dissocial PD)[14], the avoidant-self-insecure PD [15] and the BPD. Here, BPD is the best-studied PD, which is why it is given special consideration in ICD-11, also considering disorder-specific psychotherapy programs. With the dimensional PD classification system, the WHO counters the ongoing criticism of the construct of PD on the one hand and the diagnostic procedure of PD on the other hand, which have often been neglected.

A move from a categorical PD classification system towards a dimensional PD classification system was already planned for DSM-5. Nonetheless, the well-established criteria for PDs of DSM-IV remained in DSM-5 Section II whereas new, alternative criteria were placed in DSM-5 Section III for further development. In short, the newly developed "Alternative Model for Personality Disorders" (AMPD) provides a hybrid dimensional-categorical model with an evaluation of core personality functioning and five broad areas of pathological personality traits as well as identifying six specific personality disorder types.

Meanwhile, the new ICD-11 presents the first official version of a purely dimensional PD classification system with major implications for clinical practice and research. Standardized diagnostic tools are already being developed that can be used to measure PD [16] and prominent personality traits [17]. This is in line with the central aims of the new system to increase clinical utility with the assessment of severity of impairment informing clinical prognosis and treatment intensity, while specific pathological personality traits might help to inform the focus of treatment efforts[18].

CHALLENGES

Clinical challenges of personality disorders

Interpersonal dysfunction in personality disorders (PDs) are considered to be the best discriminator for diagnosis which is particularly true for borderline personality disorder (BPD)[257–259]. As a matter of fact, all effective psychotherapies share interventions to ameliorate BPD pathology by addressing interpersonal difficulties[260,261]. It is all the more interesting to note that interpersonal dysfunctions may interfere with therapeutic relationships. Hospitalized patients with BPD feel hostilely repulsed by staff members and affronted by other patients[262]. Indeed, although BPD is meanwhile considered a valid diagnosis by most clinicians, nearly half of 706 mental health clinicians in the US reported negative attitudes towards patients with BPD[263]. Thus, the diagnosis of BPD maintains a stigmatization that supports pejorative and discriminatory clinical practices such as excessive use of medication[264] despite insufficient evidence of effectiveness in BPD[265]. Correspondingly, within a 6-year period, rates of intensive polypharmacy were found to remain relatively stable over time[266]. Similarly, the management of interactive behaviors such as reactive aggression is a particular challenge when dealing with patients with antisocial personality disorder (ASPD) [267]. Currently, neither psychotherapy nor pharmacotherapy effectively addresses interpersonal difficulties in ASPD[268].

Economic challenges of personality disorders and dark personality traits

The treatment of PDs is accompanied by significant social burden and massive financial costs. For instance, the mean total costs of PDs were estimated at 11,126 euros per person annually for direct medical costs (66.5%) and indirect costs related to productivity losses in a large Dutch sample[269]. At the same time, the annual costs for untreated BPD in Germany were estimated 8.69 billion euro annually[270]. In the U.S., the costs as a result of crime attributable to individuals with the psychopathic subtype of ASPD were estimated between 245.50 billion to 1,591.57 billion US dollars[271].

At workplace, a leader's dark personality traits was associated with unfavorable outcomes, for example concerning the health of subordinates[84], counterproductive work behavior[77] or organizational citizenship behavior[80]. Applied to everyday work life, this could mean that individuals with dark traits manipulate superiors and colleagues as long as they can benefit from them, that they do not take on tasks that may be of importance for the collective but are not of individual benefit[272], or that they are also comfortable with bullying[273] as a means of eliminating their disliked opponents. Selfish individuals are assumed to be more successful than altruistic individuals within a group, but groups consisting of altruistic individuals could be more successful than groups consisting of selfish individuals[274]. Therefore, altruistic groups, led by individuals who act selfishly in the interest of the group as a whole, might be the most successful ones.

IMPAIRMENTS OF INTERPERSONAL FUNCTIONING

Borderline personality disorder

Interpersonal dysfunction has been a core feature of borderline personality disorder (BPD) from the beginning of its description[19]. Already in the first empirical study in patients with BPD, a set of criteria was suggested in which the presence of 'anaclitic relationships' was included as one of four criteria for diagnosis[20]. With further development of standardized criteria[21–23], the diagnosis of BPD entered the official classification systems DSM-III[24] and ICD-10[25]. Ever since, a pattern of instable interpersonal relationships remained a central characteristic in the diagnosis of BPD.

In the DSM-5[26], two of the nine diagnostic criteria address maladaptive interpersonal behavior in BPD, namely "frantic efforts to avoid real or imagined abandonment" and "pervasive pattern of unstable and intense interpersonal relationships". In the AMPD, two criteria have been included in BPD patients' incapacity for "intimacy" as part of interpersonal dysfunction. Additionally, the pathological personality trait "antagonism" has been proposed as a stable and consistent criterion for BPD, characterized by "hostility" that partly manifests as "anger or irritability in response to minor slights and insults". In contrast to healthy subjects, individuals with BPD are trapped in a vicious circle of negative social

interactions sustained by inter-personal hypersensitivity, affect dysregulation, and quarrelsome behavior[27]. The interpersonal difficulties of BPD patients have been related to impairments in sensing and responding to social signals. For instance, subjects with BPD tend to attribute negative emotions to neutral facial expressions[28,29]. Moreover, in these individuals, seemingly minor daily events may trigger feelings of rejection, loneliness, and failure followed by frequent, intense, and persistent aversive tension[30].

This poor quality of social interaction is thought to result from complex conditions[31]. On the one hand, affective dysregulation and impulsivity, including anger proneness, are major reasons for low interpersonal functioning in BPD [32–34]. On the other hand, abnormalities in social cognitive functions, such as impaired facial emotion recognition[29], significantly contribute to the typical interpersonal problems which are characterized by repeatedly occurring misunderstandings and experiences of being rejected and offended by others. Already in adolescents with BPD, difficulties in disengaging attention from threatening facial information during early stages of attention were found[35]. Compared to controls, adult BPD patients show lower performance when confronted with complex tasks of emotion recognition, while they show intact performance with simple tasks. For instance, BPD patients' performance was impaired in a task with integrated facial and prosodic stimuli, while they showed no impairment when instructed to recognize isolated facial or prosodic emotions[36]. Most significantly, BPD patients show a tendency to interpret ambiguous faces in a more negative or particularly more threatening way[37–39] consistent with their pervasive tendency to perceive others as malevolent[40].

There are only a few experimental studies in which participants with BPD interacted with fictitious or human confederates. When focusing on experimental studies that explicitly address interactive behavior and its modulating factors, the literature suggests an enhanced perception of and emotional distress in response to social exclusion in BPD[41,42]. Among the existing paradigms, the Cyberball paradigm has been employed the most frequently to study the perception of social exclusion in BPD. Cyberball[43] is a virtual ball-tossing game in which two to three unknown confederates include or exclude the participant from the game. Itis usually presented as an internet web page which depicts three to four animated ball-tossers standing in a circle, one of which represents the participant. When receiving the ball, the participant is asked to click on one of the others in order to throw the ball to him/her. In the inclusion condition, the participant receives as many ball passes as each of the other fictitious participants. In the exclusion condition, the participant receives either a smaller proportion or no participation at all. BPD patients showed a bias towards the perception of exclusion independently of their factual participation[44–47]. When excluded, BPD patients experienced a higher intensity of negative emotions than healthy controls[44,46,47].

Hence, the paradigm has been used for mood induction, e.g. before reward-based decision making[48].

However, it has not been explored whether the perception of social exclusion and related emotional distress have direct consequences on the interaction behavior of BPD patients. When comparing BPD patients to healthy controls, one should take into account a different baseline in hostility and negative affect towards interaction partners, which might result in more reactive aggressive behavior. To evaluate aggressive behavior of BPD patients in response to another person's prior unfair behavior, the Point Subtraction Aggression Paradigm (PSAP) [49] has been employed. The PSAP is a computer game in which participants can earn points by pressing a button 100 times. They can also subtract points from their interaction partner by pressing another button 10 times, but if they do this, these points will not be added to the their own earnings. The participants are told that their interaction partner might also take points from them. This narration blames the fictitious partner for the participants' losing points but prevents the participants from aggressive responding in order to earn money instead of earning money by pressing the other button. In two studies using the PSAP, BPD patients subtracted more points from the fictitious interaction partner than healthy controls[50,51]. Moreover, point-subtracting responses significantly correlated with self-reported hostility[50] and self-reported trait aggression, especially physical aggression[51]. Even in the condition in which the participants did not lose any points due to their fictitious partners, female and male patients with BPD and intermittent explosive disorder (BPD-IED) subtracted more points from their opponents than did healthy controls[52]. In general, BPD-IED patients pressed the subtraction button more often than healthy controls, and both groups subtracted more points from their opponents in response to pretended unfair behavior. However, in contrast to the authors' expectation, there was no group-by-condition interaction, and there were no correlations between point subtracting and clinical measures of anger or aggression in either the BPD-IED group or the HC group.

Besides the experimental settings with fictitious co-players, two studies employed real-life social interactions with human confederates to evaluate social feedback processing in BPD. In a modified "analogue" version of Cyberball, participants were asked to play cards with two attendant interaction partners [53].In contrast to the results from the virtual Cyberball paradigm, subjects with BPD did not differ from healthy controls in terms of the perception of inclusion and exclusion. However, the neural processing of social exclusion seemed to differ between BPD patients and healthy controls, with left medial prefrontal hyperactivation suggesting potential dysfunction of frontolimbic circuitry, as measured by functional near-infrared spectroscopy during the game. In a study in which participants played a well-known board game ("Monopoly"), BPD patients interacted with four healthy participants and were subsequently asked to rate themselves and one other participant on 80 character traits[54]. The rating was

conducted before and after receiving desirable and undesirable feedback from their interaction partners. Before the feedback, BPD patients rated themselves and others less favorably than did healthy controls. While healthy controls showed a positivity bias for self- and other-relevant feedback, BPD patients demonstrated a negativity bias for self-relevant feedback but not for other-relevant feedback. Especially after receiving negative feedback, BPD patients rated themselves more negatively than before the feedback. However, while both studies demonstrated alterations in either neural or behavioral feedback processing, the actual interactive behavior of BPD patients and their confederates was not described.

Given the complexity of social interaction between humans studying the interactive behavior of BPD patients, with tight experimental control on the one hand and realistic, externally valid settings on the other hand, is challenging. With regard interactive experiments in BPD, literature on chatroom paradigms and economic-exchange games will be reviewed in **chapters 2 and 3**.

Antisocial personality disorder

The clinical picture of antisocial personality disorder (ASPD) has been described with different terms in almost all societies and at almost all times[55]. In the following section, the diagnostic and research criteria for ASPD according to the DSM-5 will be explained in more detail since the diagnostics in the present work are based on DSM-5[26]. ASPD is primarily characterized by a pattern of disregard of others' rights and feelings. In the Alternative Model for Personality Disorders (AMDP), the characteristic domains of antagonism and disinhibition are central. The antagonism domain relates to immorality, grandiosity, callousness, and distrustfulness. It is a robust correlate of criminal behaviors such as theft, fraud, vandalism, and assault. Also, individuals with ASPD tend to lie, cheat, and manipulate others for personal gain. The disinhibition domain often causes individuals with ASPD to act without considering the consequences. In addition, they tend to be irritable and aggressive, which can lead them to engage in repeated fights.

Furthermore, there is the possibility of additional coding of psychopathic characteristics or other problematic personality traits. In DSM-IV field studies, 96% of individuals with psychopathy are also classified as antisocial, while 39% of individuals with ASPD are classified as psychopaths[55]. The construct of psychopathy - probably the most established subdivision of ASPD - dates back to Robert Hare (1970)[56] who developed his construct of psychopathy and its operationalization through the *Psychopathy Checklist* (PCL)[57]. The PCL includes a total of 20 items that load on two factors. The first factor—often referred to as the central aspect of psychopathy—describes selfish, insensitive, and unscrupulous personality traits and correlates highly with Narcissistic Personality Disorder and low with anxiety. Items include excessive self-esteem, lack of guilt, and lack of empathy. The second factor includes socially deviant behavior and an unstable and

antisocial lifestyle and is highly correlated with ASPD and BPD as well as reactive aggressiveness, delinquency and violence. In the context of the revised PCL in 2003[58], a four-factor model was postulated: interpersonal manipulation, callousness, unstable lifestyle, and antisocial behavior, which underlies a general psychopathy factor.

Individuals with ASPD face significant long-term impairment in family and marital relationships, occupational and residential status, and mental health[59]. As aggressive behavior is a typical manifestation, it is not surprising that the prevalence of ASPD is higher in prison than in psychiatric settings[60]. Even in prison, offenders with ASPD have difficulties in adaption because of their violent behavior[61]. Not least because of this social relevance, antisocial personality disorder is of particular importance in scientific research.

Deficits in facial emotion recognition are held responsible for aggressive behavior in antisocial personality disorder (ASPD)[62] or rather its subtype psychopathy[63]. For example, psychopathic individuals showed selective deficits in the recognition of fear[64], and psychopathic traits were found to be associated with reduced fixations to the eye region of fearful faces[65]. However, most experimental studies which drew conclusions on ASPD have been done in either adolescents with antisocial behavior[66], healthy individuals with symptoms relevant to ASPD, or forensic populations with psychopathy[67]. There is only one experimental interactive study which included six male individuals with ASPD and no control subjects and measured aggressive responding in the PSAP [68]. The main objective of this study was to measure acute intranasal oxytocin dose effects on monetary-reinforced responding in ASPD and found no effects. Hence, the lack of research in this area indicates that the experimental examination of interpersonal dysfunctioning in ASPD is novel. Therefore, a well-studied paradigm of facial emotion recognition is used in **chapter 4** in order to study aspects of social cognition as one basis of interpersonal behavior in this relevant but underexplored psychiatric disorder.

Dark personality traits

With a limited number of participants with ASPD in experimental studies by the very nature of participants having ASPD[67], interactive behavior has been studied in individuals with personality traits relevant to the symptomatology of ASPD. Interestingly, the other side of the coin is the overrepresentation of individuals with antisocial personality traits in leadership positions[69–71]. Thus, antisocial behavioral strategies have even been considered to be functional by some authors, as the individual benefits at the expense of the social environment [72]. The Dark Triad (DT) is the most frequently studied cluster of "dark" personality traits in relation to leadership[70]. The DT personality construct was constructed as a cluster of three interrelated but distinct personality traits[73]: Machiavellianism, (subclinical) narcissism, and (subclinical) psychopathy. All three traits are considered to be

dimensional in nature. Of the three Dark Triad traits, Machiavellianism, also called the *manipulative personality*, is characterized by superficial social relationships with little affective involvement, deviations from normative and ethical behavior, and a pragmatic (rather than an ideological) orientation[74]. The construct of subclinical narcissism dates back to Raskin and Hall's (1979) *Narcissistic Personality Inventory*[75] including the facets of grandiosity, entitlement, dominance, and superiority. Subclinical psychopathy emerged from Hare's (1985) *Self-Report Psychopathy Scale*[76] and is related to impulsive and thrill-seeking behavior in combination with low empathy and anxiety.

Studies have demonstrated that a leader's dark personality traits are related to various undesirable outcomes both at individual and organizational levels. Such associations have been found in relation to increased counterproductive work behavior[77], lower integrity[78], reduced transformational leadership[79], and deteriorated organizational citizenship behavior[80] displayed by the leader. With respect to the subordinates of dark leaders, their well-being[81], job satisfaction[82], and organizational commitment[83] are reduced. Accordingly, more frequent burnout cases[84] and turnover intentions were reported[83]. Most previous studies have only revealed descriptive relationships between the DT and behavioral outcomes. Studies that look for explanations of the nature of the relationships between the DT and the rather uncooperative and antisocial behaviors associated with the DT in controlled experiments are scarce.

Previous studies on uncooperative behavior in individuals with DT traits have employed the Prisoner's Dilemma (PD), in which two participants have to choose between cooperation and defection. The cooperative choice maximizes joint profit while the non-cooperative choice maximizes individual profit[85]. The DT construct as a whole[86], psychopathy[87–89], as well as Machiavellianism and psychopathy combined[90], have been linked to less cooperative game playing strategies. With one exception[88], all studies used computer-simulated interaction partners and deceived the participants about their true opponents. These studies also examined the effects of framing [86], diverse effects of bargaining under distinct types of social interactions[88], the significance of different subscales of psychopathy on a sample of criminal psychopaths[87], and the role of impulsivity[90]. Given the clear evidence of the potential harm resulting from dark leaders in positions of power, **chapter 5** examines individuals' Dark Triad characteristics in three well-described economic-exchange games without using deception, in which participants have to consider the consequences of one's own decisions for one's own progress, the well-being of others and delay discounting.

NEUROBIOLOGY OF IMPAIRMENTS OF INTERPERSONAL FUNCTIONING ON THE EXAMPLE OF FACIAL EMOTION RECOGNITION

Neural processing

There are a number of theories on the etiology of BPD and ASPD which include genetic, neurobiological and environmental models[91]. More recent studies have also looked at specific neurobiological factors, such as brain circuits[92]. Alterations in the capacity to form relationships in an empathic, cooperative way are common to BPD and ASPD, particularly the psychopathic subtype. Apparently, the nature of social dysfunction among these disorders is heterogeneous. Recent progress in social neuroscience has contributed to understanding the differential brain mechanisms mediating specific patterns of interpersonal dysfunctioning. For instance, impairments in neural processing of facial emotion recognition have been described in both borderline personality disorder and antisocial personality disorder.

In BPD, reduced thresholds for anger detection in predominantly happy faces were associated with higher occipital P100 amplitudes reflecting initial visual processes and with lower amplitudes in phases representing structural (N170) and categorical (P300) facial processing[93]. However, P300 amplitudes were increased for highly angry faces in BPD, whereas in healthy volunteers this component was higher in response to both, angry and happy faces. In line with this, BPD patients misclassified briefly presented affective faces as angry and showed faster initial saccades towards angry eyes, but shorter fixation times on angry eye regions compared to healthy controls. This early sensory bias for interpersonal threat might be driven by the amygdala which is supported by a brain imaging study detecting enhanced amygdala activation to overt fearful faces in BPD together with hyperactivation in prefrontal cortex (PFC) regions involved in emotion regulation[94]. The latter finding may suggest that patients try to control emotions – albeit unsuccessfully, probably due to a deficient functional interplay within the prefronto-amygdala circuit[95].

In ASPD, poor recognition of fearful faces has recently been linked to trait callousness and associated with reduced N170 amplitudes over the fusiform face area (FFA) and surrounding regions[96]. Reduced activations in the FFA and other face processing areas, such as the superior temporal sulcus (STS), were also found in psychopathic criminal offenders to dynamic presentations of fearful, but also sad, happy and painful facial expressions[97] suggesting a pervasive deficit in facial emotion processing beyond fear. Notably, psychopaths with larger volumes of the prefrontal and cingulate cortex, anterior insula, somatosensory cortex, and the posterior lobe of the cerebellum exhibited better facial emotion recognition[98].

The role of oxytocin

Impairments in social cognition may be linked to alterations in neuropeptide systems, such as the oxytocin system[99]. Oxytocin is a neuropeptide that is central to various aspects of human interpersonal behavior and social cognition, i.e., increase cooperation, empathy and trust[100].

There is first evidence that the intranasal application of oxytocin may decrease threat hypersensitivity in BPD patients[101]. Using an eye-tracking methodology, BPD patients were shown to exhibit more and faster initial fixation changes to the eyes of angry faces, and this behavioral pattern was associated with increased posterior amygdala activation compared with the control group. The authors showed that these abnormal behavioral and neural patterns were normalized after oxytocin administration. However, oxytocin has been also shown to hinder trust and cooperation in a small sample of individuals with BPD[102] (*N* = 14). In a small sample of individuals with ASPD (*N* = 6), there was no specific effect of OT on aggressive responding in the PSAP[68].

The potential application in personality disorders of oxytocin have not yet been explored. The current guidance on the treatment of BPD and ASPD highlights that the evidence base for both pharmacological and psychological interventions is limited[103].

AIMS AND SCOPE OF THIS THESIS

Social interaction is complex and there are many ways in which social interaction can be impaired. So far, various aspects of social interaction but especially social cognition have been studied. Therefore, the main objective of this thesis is to investigate social interaction of individuals with antisocial behaviors in ecologically valid experiments.

Since borderline personality disorder (BPD) is a prime example of a mental illness with severe impairments in social interaction, we first focus on new interactive experiments in BPD. Specifically, **chapter 2** examines emotional responses to receiving peer feedback on opinions in a chatroom paradigm. Further, **chapter 3** evaluates social decision-making and partner preference in economic-exchange games.

Second, we investigate individuals who are less present in the clinical context but whose severe impairments in social interaction have a serious societal impact. In detail, **chapter 4** investigates facial emotion recognition in young adults with antisocial personality disorder (ASPD) and its modulation by oxytocin. Moreover, **chapter 5** explores the relationship between dark personality traits, sex, and socio-economic decision-making connected to leadership emergence.

Finally, **chapter 6** discusses the results of these studies in light of current clinical and societal implications, and future directions of research are proposed.

CHAPTER 2

2

Emotional responses to receiving peer feedback on opinions in borderline personality disorder

Haang Jeung, Stephan Walther, Christoph W. Korn, Katja Bertsch, Sabine C. Herpertz

Personality Disorders: Theory, Research, and Treatment, 2018

ABSTRACT

Although emotional reactivity to social rejection has been examined in patients with borderline personality disorder (BPD) in several studies, the effects of other aspects of social feedback, such as evaluation of one's opinions that concern self-esteem, have not been addressed yet. The objective of this study was to examine emotional responses of BPD patients after exchanging personal opinions in a new, ecologically valid virtual peer interaction paradigm ("chatroom paradigm"). In this paradigm, 21 BPD patients and 21 healthy controls received peer feedback on their own statements and rated the intensity of their own emotional responses (happiness, sadness, anger, and shame) and the self or other affirmation in response to agreement, disagreement, and neutral statements. Across all social feedback conditions, BPD patients reported more intense negative emotions and less happiness than healthy controls. While healthy controls showed a "positivity bias" for any type of social feedback, the emotional responses of BPD patients corresponded to the valence of the feedback; that is, they were happiest after positive than after neutral feedback and least happy after negative feedback. Disagreement resulted in more intense anger and less other affirmation in both groups but only BPD patients also experienced higher shame in this condition. This is the first study to assess emotional responses to social feedback in an ecologically valid chatroom paradigm. Our findings underline that more negative emotional reactions in everyday interactions play a central part in interpersonal difficulties of patients with BPD.

INTRODUCTION

Difficulties in interpersonal relationships are one of the most stable and deliberating symptoms of borderline personality disorder (BPD)[104]. BPD patients report that they experience less social support, more conflicts, and less integration in their social networks[105]. Real-life interpersonal interactions trigger more negative emotional responses in BPD[27,106,107].

Previous studies on social feedback in BPD have addressed the preference for affiliating with the participant versus preferring a fictitious other which target patients' sensitivity to threats to belonging[108]. Most experimental studies focused on social rejection using the Cyberball paradigm, a virtual ball-tossing game, which induces more intense negative emotions – especially elevated anger – in BPD patients than in healthy volunteers (e.g.[46]). These increased negative emotional responses of BPD patients may be explained by a rejection hypersensitivity, that is, anxious expectation and perception of and overreaction to rejection cues[109].

There have only been few studies dealing with other social needs of human beings, such as self-esteem, which respond to social evaluations, such as providing social feedback on personality characteristics or task performances, as they frequently occur in everyday social life. For instance, more intense feelings of shame were found in individuals with high borderline personality features after receiving a negative evaluation for their task performance and more intense anger to a negative evaluation of personal characteristics[110]. Additionally, socially negative feedback reinforced negative self-evaluation in BPD[54]. A recent study however found that in real-life, only perceived rejection but not criticism predicts negative emotional responses, such as anger and shame in BPD[107].

One problem of previous experimental studies examining emotional responses in BPD patients is the use of standardized or artificial interactive stimuli (Cyberball), making it difficult to translate their findings to real-life interpersonal interactions. To mimic live interaction in a more ecologically valid way, chatroom paradigms have been developed and tested in the last decade[111]. To overcome problems associated with rather narrowly defined social feedback associated with "liking" and "disliking" in classical Chatroom paradigms, we developed and tested a new, close to real-life interaction paradigm to examine emotional responses of BPD patients when confronted with the challenge to express personal opinions and receive peer feedback on these opinions.

Present study

Since BPD symptoms have been associated with less positive emotions[112], vulnerability to anger[113], shame, and low self-esteem[114], we examined these specific emotional responses together with self or other affirmation, that is, commitment to one's own opinion and acceptance of another person's opinion, during a chatroom conversation. First, we hypothesized that female BPD patients

would report more intense anger and shame in response to disagreement and less happiness in response to agreement than healthy women. Second, we expected BPD patients but not healthy controls to show less self-affirmation after disagreement than after agreement, whereas the levels of other affirmation would be similar to healthy controls.

METHOD

Participants and recruitment

A total of 21 female BPD patients (BPD; M_{age} = 24.3, *SD* = 3.5 years) and 21 female healthy controls (HCs; M_{age} = 23.9, *SD* = 3.3 years) took part in the study. BPD patients were outpatients of the Department of General Psychiatry, University Hospital Heidelberg, healthy controls were recruited from the community through online advertisement and flyers. The groups were matched for age and educational background.

Due to the high comorbidity rates with other psychiatric disorders[115], we did not exclude concurrent diagnoses in BPD patients. General exclusion criteria comprised neurological diseases, history of head trauma, and any severe medical condition. Additionally, only BPD patients without current alcohol or drug dependence, acute and chronic psychotic disorders, or bipolar disorders were excluded. Only HCs without any lifetime psychiatric disorder including BPD were enrolled. The study was approved by the local Ethics Committee of the University of Heidelberg. Participants gave written informed consent and were debriefed carefully about deception and were remunerated
after their participation.

Procedure

The *Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Axis I Disorders*[116] and the *International Personality Disorders Examination, BPD section*[117] were conducted by a trained psychiatrist (H.J.).

Chatroom task

Participants were told that the aim of the study was to examine the exchange of attitudes toward different values through internet-based communication and that, to this end, they would interact with an unfamiliar female peer in a chatroom. In fact, the fictitious interaction partner was a computer program that generated either positive or negative feedback on the basis of a randomized algorithm. Before the task started, a photograph was taken of each participant to be integrated into the interface of the chat program. Then, the participant took part in a fictitious lottery that determined her role as either questioner or responder. In

fact, she was always appointed as responder. To foster the credibility of the predefined questions, each participant was told that the other person would primarily ask her own questions but was allowed to choose topics from a list. The task had two phases:

Answering questions. In the first phase, participants were asked to type their individual answers to 15 open questions about their person, for example, zodiac sign (control condition) and 30 open questions about their opinions on values, for example, vegetarianism or religion (experimental condition) adopted from the European Values Study (http://www.europeanvaluesstudy.eu). After each question about their opinion, participants were asked to rate on a scale from 1 (*not at all*) to 4 (*very much*) how certain they were of their opinion. In total, the first phase took approximately 30 min.

Receiving social feedback. *Figure 1* shows a schematic illustration of the second phase of the task: The participants received either positive or negative feedback from their fictitious peer. One question, its answer, and corresponding feedback formed one unit, which was presented in a block consisting of three units of the same feedback entity. Each unit was presented for 3 s, with a jittered time of 2 s between each unit. In total, 10 blocks were presented, with five blocks of negative feedback (e.g., "I disagree") and five blocks of positive feedback (e.g., "I agree"). After each block, participants rated the intensity of their emotions (happiness, sadness, anger, and shame) on a scale from 1 (*not at all*) to 4 (*very much*), as well as commitment to their previous opinion (as a measure of self-affirmation) and acceptance of the other person's opinion (as a measure of other affirmation) on a scale from 1 (*not at all*) to 4 (*very much*). The order of the emotion ratings was randomized, and ratings of self and other affirmation always appeared after the emotion ratings. In addition, five blocks were presented in which questions and answers about the personal characteristics of the participant and the other person were shown in groups of three (e.g., "My zodiac sign is Aries"). In total, the second phase lasted approximately 15 min.

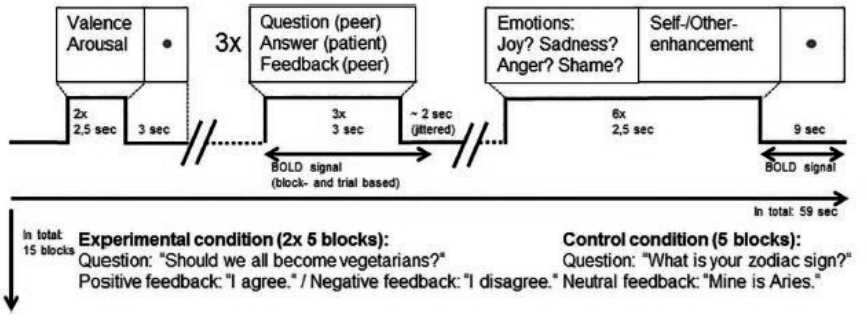


Figure 1. Chatroom paradigm (block design)

Baseline and final assessments

When the chat program started, the photograph of the fictitious partner was shown, and each participant was asked to rate her liking of the interaction partner on a visual scale from 1 (*thumbs down*) to 4 (*thumbs up*) at the beginning and end of the experiment.

After the first phase, the 30 answers on opinions were matched according to certainty ratings and randomly grouped into threes, generating 10 feedback blocks with equal levels of certainty. Within a block of three questions/answers, there was either positive or negative feedback. Before the second phase, participants were told that a computer grouped question/answer blocks according to feedback. In the second phase, each participant indicated on a modified 4-point Self-Assessment Manikin scale how pleasant and aroused she felt before each feedback block. A fixation cross appeared nine seconds before and three seconds after each Self-Assessment Manikin rating. Following task completion, participants were asked, using open questions, whether they noticed any peculiarities about the experiment. Based on their comments, the experimenter (H.J.) rated the credibility of the setting (1 = *no peculiarities found*; 2 = *doubts about the partner's existence*; 3 = *other comments*).

Data analysis

For all analyses, the significance level was set at an alpha of .05. For demographic and clinical characteristics and baseline and final assessments, categorical data were analyzed with nonparametric statistics (χ^2 tests). Other comparisons were conducted with *t* tests for independent samples and one-way analyses of variance (ANOVAs). For analysis of the task, we mainly used repeated measures
ANOVAs with the between-subjects factor group (BPD/HC) and the within-subject factor feedback (either negative/positive/neutral for emotional intensity ratings or negative/positive for self- or other-affirmation ratings). In the case of violation of the assumption of sphericity, Huynh–Feldt correction was applied. Dunn's multiple comparison procedure including Bonferroni correction for multiple testing was used as post hoc tests.

RESULTS

Participant characteristics

The BPD and HC groups did not differ in age, years of education, nationality, or relationship status, but significantly more HCs reported current employment than BPD patients.

Hypothesis 1: Emotional Intensity After Social Feedback

First, we checked whether the groups differed in terms of emotional intensities and found a significant group by emotion interaction, $F(1.56, 62.630) = 21.03$, $p = .001$, $\eta^2 = .35$, indicating that, indeed, different emotional responses were found in BPD patients and HCs. Most interestingly, this effect was qualified by a significant Group X Emotion X Feedback interaction, $F(3.92, 156.7) = 3.62$, $p = .05$, $\eta^2 = .08$, confirming our a priori hypotheses. For illustration, differences in mean intensities of emotional responses are shown in Figure 2.

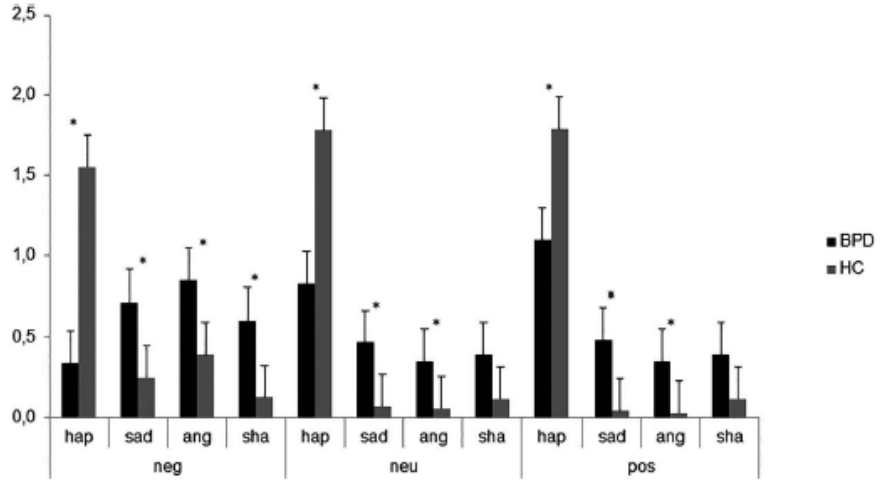


Figure 2. Mean emotional intensity rating after negative, neutral, and positive feedback in borderline personality disorder (BPD) patients and healthy controls (HCs).
*= significant at $p < .05$

Anger. In all pairwise comparisons, the intensity of anger was higher in the BPD group than in the HC group (*p's* = .001 for negative and positive feedback conditions, and $p = .05$ for the neutral condition). For each group, the intensity of anger was significantly higher after negative feedback than after positive or neutral feedback (all *p's* = .001), respectively. All other comparisons were nonsignificant.

Shame. Only after negative feedback the intensity of shame was higher in the BPD group than in the HC group ($p = .001$). After positive or neutral feedback, the intensity of shame did not differ between the two groups.

Happiness. Interestingly, the post hoc tests of the three-way interaction (Group X Emotion X Feedback) showed that the intensity of happiness remained at the same level across all conditions in the HC group, whereas it changed as a function of the feedback condition in the BPD group. In the latter group, intensity

of happiness was highest after positive feedback and lowest after negative feedback (all *p's* = .001).

Sadness. Sadness was rated significantly higher in the BPD group than in the HC group (all *p's* = .001). Within groups, intensity of sadness was on a similar level within and across conditions.

Hypothesis 2: Self or other affirmation after social feedback

Here, neither the two-way nor the three-way interaction with the between-subjects factor group reached significance.

Credibility of the Manipulation

The majority of participants (72.4% of BPD patients and 66.6% of HCs) did not report any doubts about the credibility of the setting. We did not exclude any data in this respect, as there were no statistically significant differences between the two groups.

Partner liking and certainty about self-statements

Pre- and post-task ratings were compared in a 2 X 2 ANOVA with the factors group (BPD or HC) and measurement time (pre- and post-task). According to this, there were no group differences in the liking of the interaction partner (p = .05) for main effect of group and all interactions including the factor group but a general decline in the liking from pre- to post-task ratings independent of group (main effect of time: $F(1, 40)$ = 15,279, p = .001, η^2 =.276. Furthermore, BPD patients and HCs did not differ in certainty about self-statements before they received social feedback.

Emotional recovery

We checked valence and arousal before we started each new block. Here, the 2 (group: BPD/HC) X 2 (emotional state: valence/arousal) X 3 (feedback: negative/positive/neutral) ANOVA showed no three-way interaction. Most importantly, there was no main effect of feedback condition. However, there was a significant group by emotional state interaction. In the pairwise post hoc tests, the arousal was significantly higher in the BPD group than in the HC group (p = .001).

DISCUSSION

The aim of the present study was to assess how positive or negative feedback in the form of agreement and disagreement of participants' opinions issues of self-esteem among peers affect emotional reactivity and self or other affirmation in BPD. Across all feedback conditions, BPD patients exhibited more intense negative

emotions and less happiness in response to social feedback than HCs. Moreover, BPD patients reported more intense anger and shame in response to disagreement and less intense happiness in response to agreement than HCs. This emotional response pattern is consistent with findings of intense negative emotions in response to social rejection[46] and avoidance of positive emotions[118]. Contrary to our expectations, self-affirmation remained stable across feedback conditions in both groups. This was somewhat unexpected, as in a prior study, BPD patients rated themselves more negatively after negative self-relevant feedback but not after other relevant feedback, in contrast to healthy controls[54]. This could indicate a distinction between different types of feedback, for example, on personality characteristics as opposed to values/attitudes.

Interestingly, HCs showed a "positivity bias" for any type of feedback, whereas BPD patients responded in accordance with the valence of feedback: They were happiest after positive feedback, less happy after neutral feedback, and the least happy after negative feedback. Similarly, different types of positivity biases[119] have been repeatedly found in healthy participants[120].

When interpreting the specific emotional responses in detail, we would first like to notice the similarity of enhanced anger of BPD patients and HC to the negative versus neutral or positive feedback in the current study and social exclusion versus inclusion in previous Cyberball paradigms[46]. However, only BPD patients reported also more intense feelings of shame after negative feedback, a finding that is in line with previously reported emotional responses to feedback about poor task performance[110]. Furthermore, sadness was generally higher in BPD patients than in HCs independent of feedback condition, which points to overall elevated feelings of sadness in BPD patients. Again, this is in line with higher levels of sadness in BPD patients before and after Cyberball[46].

The differential pattern of these emotional responses to social feedback in the current chatroom situation might be explained by differences related to emotions themselves: While happiness and sadness are outcome-dependent emotional responses to success and failure, shame and anger are emotional responses with attributions to self and others[121].

LIMITATIONS AND FUTURE DIRECTIONS

This is the first study assessing emotional responses and self or other affirmation to social feedback in a close-to-reality chatroom paradigm. One limitation is the sample size of 21 participants per group. There are growing concerns regarding the replication of laboratory findings in clinical psychological research[122]. Despite an a priori power analysis that was based on results of a pilot study and according to which n = 14 participants/group are sufficient to detect a moderate effect (d_{Cohen} = .7) with a statistical power of 1-ß > .80, further replications with the current and

similar paradigms are needed before strong conclusions can be drawn. As no differences in ratings of self or other affirmation were found between BPD patients and HCs, the questions about self or other affirmation in our paradigm might not have been sufficiently discriminatory to detect differences between the two groups. Alternatively, the values/attitudes have not been important enough to influence self or other affirmation, which points to a potential dissociation between different types of feedback (i.e., on personality characteristics and values/attitudes).

In future functional imaging studies, we would like to disentangle the neural facets of emotional responses in complex interpersonal situations. To enhance power for such functional imaging studies, we designed our chatroom paradigm using a so-called block design. In order to develop intervention strategies, further studies should examine active processes to modulate one's emotion and self-esteem in response to an emotional stimulus[123].

CONCLUSION

Our findings suggest a differential emotional response of patients with BPD to social feedback in forms of agreement/disagreement to opinions on values that concerns self-esteem. Negative social feedback resulted in higher anger and less other affirmation in both groups but a heightened reactivity of shame toward the self in response to negative social feedback in BPD only. In BPD, the level of happiness was graduated in correspondence with the type of feedback. In contrast, HCs showed a positivity bias. When translating our findings to clinical practice, therapists should call their attention to the regulation of shame in the therapeutic process of patients with BPD.

CHAPTER 3

3

Consider others better than yourself: Social decision-making and partner preference in borderline personality disorder

Haang Jeung, Martin Vollmann, Sabine C. Herpertz, Christiane Schwieren

Journal of Behavior Therapy and Experimental Psychiatry, 2018

ABSTRACT

Patients with Borderline Personality Disorder (BPD) suffer from interpersonal difficulties. They have been shown to be distrustful and yet involved in abusive relationships. In this study, we want to examine whether the perception of fairness and partner preference are altered in BPD. We employed a coalition formation game in which a participant can choose whether to interact in dyads or triads, thus exclusion or inclusion of a third potential interaction partner. Furthermore, triads get a higher endowment, such that dyads are not only unfair to one partner, but also economically inefficient, as the participant reduces the overall amount of money available for distribution. Subsequently, we compared how participants predicted another person's game strategy (inclusive, exclusive, or mixed) and rated its fairness, and which partner the participant would select. The majority of the BPD group (n = 26) as well as of the healthy group (n = 29) preferred triads over dyads and offered a near-to-equal split to their interaction partners in the first two rounds. In contrast to the healthy group, the BPD group did not show a drop of the average level of investment in the final round. In both groups, the inclusive strategy was perceived as the fairest strategy. Most interestingly, despite a similar perception of fairness, half of the BPD group preferred an interaction partner with an exclusive or mixed strategy while the majority of the HC group would choose an interaction partner with an inclusive strategy. This is a preliminary study which needs further replications before strong conclusions can be drawn. Our study demonstrates no differences in fairness perception but an alteration in partner preference of patients with BPD which might contribute to unfavorable partner choices and impairments of interpersonal functioning in BPD.

INTRODUCTION

Borderline Personality Disorder (BPD) is characterized by severe and persistent impairment in interpersonal functioning[124]. Compared to healthy controls, patients with BPD report less social support, more conflicts, and less integration in their social networks[31,41,105]. Some of the interpersonal difficulties experienced by patients with BPD might result from the choices that patients with BPD make in relationships. For instance, they prefer few but close and tense relationships[125]. With regard to partner preference, female patients with BPD tend to engage in romantic relationships with men who, in particular, have antisocial personality disorder[126] which in itself is a risk for intimate partner abuse[127]. However, there are only a few experimental studies which examine the interpersonal choices made by patients with BPD.

In previous economic-exchange studies, patients with BPD seem to act mostly "rationally" in their own self-interest and independently from social signals[128]. Their motives for non-cooperation could be lower trust in others[129], lower trustworthiness[130], and negative reciprocity in the sense of "tit for tat" type of response[130,131]. However, trust and reciprocity are not the only social preferences that shape social decision-making. In one-shot encounters, strategic behavior concerning reciprocity should not matter[132]. Nevertheless, fairness motives significantly affect human behavior independent of the strategic situation[133]. Fehr & Schmidt[133] describe "inequity averse" individuals as making decisions so as to minimize inequity in outcomes. Previously, we have proposed that individuals with BPD show less inequity aversion than healthy individuals when it comes to the acceptance of unfair offers from others[128].

Interestingly, there are conflicting possibilities of interpreting the patients' sense for fairness. On the one hand, patients with BPD engage in altruistic punishment, i.e., they punish fairness violations of others even at their own cost, just as well as healthy controls[134]. On the other hand, they accept unfair offers by others more frequently than did healthy controls[135]. This appears to be contradictory at first sight but it can be explained by the unequal treatment of *another person* or *oneself*. Most of them having experienced abuse and neglect in childhood, patients with BPD have been described to react strongly emotionally, mainly angrily, with the urge to defend the right of *others* when they observe injustice[136]. Three previous studies have examined how injustice and unfairness towards others and themselves may affect social interactions in BPD[137–139]. Patients with BPD did not only report higher justice sensitivity as compared with healthy controls but were also more willing to behave solidary in a lottery game[138]. At the same time, patients with BPD might perceive unfair treatment to the detriment of *themselves* as deserved as they expected a more negative outcome and anticipated less positive emotions in case of a positive outcome for themselves. Similarly, patients with BPD rejected higher rates of fair offers and reported more

anger and less happiness than healthy controls after fair offers in a modified ultimatum game[137]. Moreover, 50% of patients with BPD rejected the total endowment offered by a proposer whereas only 8% healthy controls did so[139]. In parallel, patients with BPD expect unfair treatment by others as they have repeatedly shown a bias towards the perception of exclusion, independently of their factual participation, and a higher intensity of negative emotions after exclusion in several Cyberball paradigms[46,47]. Yet, interaction is bilateral in nature. Up to now, it has not been studied whether patients with BPD make fair offers to their interaction partners once they are in control of inclusion and exclusion without future reciprocity concerns.

OVERVIEW AND HYPOTHESES

It has been described before that individuals with BPD do not cooperate with others because they distrust them. In this study, we want to examine fairness in BPD independent of reciprocity. Therefore, we applied an economic-exchange game that has been previously described to study social exclusion by non-cooperative coalition formation in non-psychiatric subjects[140]. More precisely, subjects in the game can choose whether to interact in dyads or triads, where the dyads are characterized by excluding a third potential interaction partner and also depriving this person of the possibility to earn money from the game. Furthermore, triads get a higher amount of money as endowment, such that dyads are not only unfair to one partner, but also economically inefficient, as the subjects reduce the overall amount of money available for distribution.[b]

Hypothesis 1. Given their difficulties with interpersonal trust and preference for a few close relationships, we hypothesized that patients with BPD, in contrast to healthy controls, preferred dyads over triads, and thus, the social exclusion of the third person.

Hypothesis 2. At the same time, due to their higher justice sensitivity towards others, patients with BPD would offer their interaction partner a fair or even higher split of the endowment in contrast to healthy controls.

Hypothesis 3. As the literature indicates no differences between the two groups in judgment, patients with BPD and healthy controls would rate, without differences, triads to be fairer than dyads.

Hypothesis 4. *W*hen it comes to partner preference, however, due to the tendency to engage in unfair, abusive relationships, patients with BPD but not healthy controls would choose an interaction partner with a preference for dyads.

[b] When talking about inefficiency we refer to it in a monetary sense, reflecting forgone resources[140]. In our case less money can be distributed in the dyad compared to the triad.

MATERIALS AND METHODS

Participants and recruitment

Twenty-six women with borderline personality disorder (BPD) and twenty-nine female healthy controls (HCs) matched for age (18–40 years) and educational background took part in the study. Demographic and clinical characteristics are shown in *Table 1a*. Patients with BPD were outpatients of the Department of General Psychiatry, University Hospital Heidelberg, healthy controls were recruited from the community through online advertisement and flyers. BPD exclusion criteria comprised neurological diseases, history of head trauma, current alcohol or drug dependence, acute and chronic psychotic disorders, bipolar disorders, a history of illicit drug use in the previous 2 months, alcohol dependence or abuse for the previous 2 months as well as any medical condition that may affect central nervous system functioning. Only HCs without any lifetime psychiatric disorder including BPD and without taking any psychotropic medication were enrolled.

Further, we employed 120 co-players who came from our standard pool of student subjects. The experiment was organized and student subjects were recruited with the software *hroot*[141], an online recruitment system for economic experiments.

Clinical assessment

All patients and healthy controls underwent clinical assessment with the *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)*[116] and *International Personality Disorders Examination, BPD section (IPDE-BPD)*[117]. All interviews were conducted by the principal investigator who is a trained senior psychiatrist. In all participants and student subjects, we collected commonly used measures of symptom severity, namely the *Brief Symptom Inventory* (BSI[142]) and *Borderline Symptom List* (BSL-23[143]), in order to characterize our sample.

Procedure

After a complete description of the study to the patients and controls, written informed consent was obtained. The study was approved by the local Ethics Committee of the University of Heidelberg.

On the day of the experiment, we invited three focal participants (either patients with BPD or healthy controls) and six student subjects who were not aware that this study also involved patients. Additionally, we invited one extra student subject in case that one participant did not show up. If the focal participant did not show up, this student subject participated instead, but her data was not included into the analysis. A total number of nine participants was required for each session to fulfill the randomization conditions. All participants and student subjects were asked to report their medication as prescribed. The majority of the

patients with BPD disclosed to take psychotropic medication (69.3%). All participants and student subjects received a show-up fee of € 4.00 plus earnings from the experiment which is described below.

Tasks

The experiment was conducted in the laboratory of the Alfred Weber Institute of Economics, University of Heidelberg, and consisted of two tasks.

Task 1: Coalition formation. As laboratory task, we adapted a non-cooperative three-person coalition formation game with an ultimatum bargaining stage[140] which was conducted through the Zurich Toolbox for Ready-made Economic Experiments (z-Tree)[144], version 3.4.7. Each game consisted of three rounds. In order to prevent strategic and reciprocal behavior, one participant and two students were randomly assigned to a group of three for each round. The randomization was done in a way that a player was never matched twice with the same partner. The three players involved were called proposer, responder 1, and responder 2. Patients and healthy controls were always proposers and students were always responders. Further, the players did not know the identities of their fellow players. The sequence of the play was the following (see also Fig. 1): A proposer could choose between forming a dyad (two-person coalition), excluding one of the responders[c] or a triad (three-person coalition), including all players. A dyad had a value, i.e., an endowment, of 2100 points to be split in intervals of ten between the members of the dyad. A triad had a value of 3000 points to be divided between all three of them. The decision for the dyad therefore reflects a monetary efficiency loss of 30% (or 900 points) compared to the triad. The proposer had to divide the coalition value between herself and the chosen responder(s). If she chose the triad, she could choose how to split the money between herself and the two responders, such that both responders received the same amount of points and she could keep the rest. If she chose a dyad, she would propose one split only with the chosen responder and keep the rest, while the third person did not get any points. If a responder was chosen as a member of either the dyad or the triad, she could decide whether to accept or reject the proposal. If the chosen responder(s) accepted the proposal, all players would receive their shares. Otherwise, nobody earned anything. If a potential responder was not chosen, she could not influence the outcome and had a zero payoff from that part of the game. The exchange rate from points to Euro was 500 points = €1.

[c] If the proposer chose the dyad one random responder was excluded from the coalition.

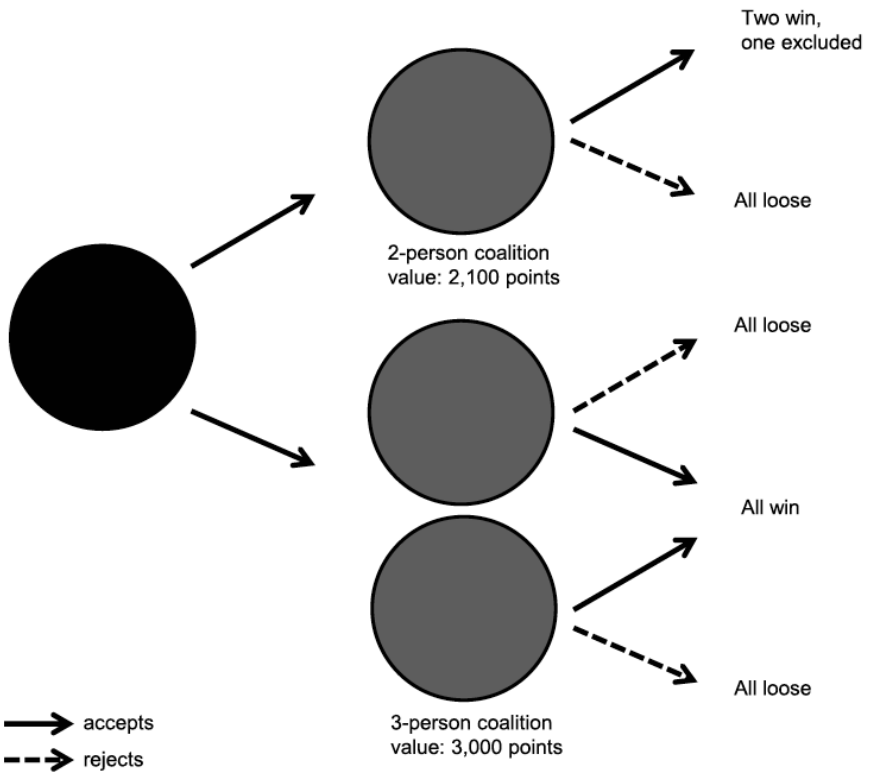


Figure 3. Coalition formation game. In each round, proposers (either individuals with borderline personality disorder or healthy controls) were randomly assigned to two interaction partners (students). Proposers could choose to form a dyad (two-person coalition) or triad (three-person coalition) with one or both interaction partners. The dyad had a coalition value of 2100 points (endowment), the triad had a coalition value of 3000 points (endowment). Interaction partners who were excluded from the dyad were observers and could not participate in this round as a responder. Proposers could offer responders a split of the endowment. If (all) responders accepted the offer, the points were divided amongst proposers and responders as proposed. If (one of the) responders rejected the offer, no one received any points.

Task 2: Judgment, fairness ratings, and partner preference. In the second task, all participants watched the possible game histories of three separate proposer types in the same game that they had played before. It was explicitly explained that they should review the three distinct strategies carefully, one by one. After seeing the coalition decision, the participant had to indicate on a scale from 0 to 100% how high she estimated the chance that the proposer she was observing would choose a triad in the next round. . Each game consisted of nine rounds, hence, the number of guesses was eight for each proposer type. In one

condition ("inclusive strategy"), the proposer always chose triads, in another condition ("exclusive strategy"), the proposer always chose dyads. Additionally, there was a third condition in which the proposer randomly chose triads or dyads ("mixed strategy"). The strategies were presented in a randomized order.

After having watched all game histories, participants were asked to rate the fairness of each proposer type on a scale from 0 (= not at all) to 4 (= very much). Finally, they had to select the proposer type they would prefer to play with in a game. To prevent participants from hedging[d] [145] their guesses and choice did not have any payoff consequences.

Statistical analyses

The data were analyzed with SPSS (Version 24; SPSS Inc., Cary, NS). For all analyses, five percent was chosen as the level of statistical significance. Categorical data were analyzed with nonparametric statistics (χ^2 tests). All other comparisons were conducted with t-tests for independent samples and one-way analyses of variance (ANOVAs).

RESULTS

Participants' characteristics

Means and standard deviations for descriptive statistics and all self-report measures are presented in *Table 1a* for patients with BPD and HCs and in *Table 1b* for student subjects, respectively. Initially, 30 individuals with ASPD and 30 HCs were enrolled in the study. Five participants dropped out of the study not showing up (four BPD and one HC). Their participation was replaced by student subjects whose data were not analyzed further.

[d] If guesses would be incentivized they would become part of the payoff relevant action space, which would give the participants the possibility to use their stated beliefs to offset the risk of adverse outcomes in the rest of the experiment (Blanco et al., 2010, p. 413).

Table 1. Demographic and clinical characteristics of participants and interaction partners

a. Demographic and clinical characteristics of proposers

	HCs (n = 29)		BPD (n = 26)				
	M	SD	*M*	SD	*t*	*p*	
Age (in years)	23.3	4.6	24.6	3.2	-1.2	.2	n.s.
Education (in years)	11.1	1.1	11.2	1.1	-0.2	.9	n.s.
BSI GSI	.2	.2	1.8	.7	-911	<.01	**
BSL-23 Score	.2	.2	1.9	.9	-9.1	<.01	**
	n	%	n	%	c^2	*p*	
Currently employed	26	90	12	46	12.2	<.01	**
In a relationship	18	62	13	50	.81	.4	n.s.
Children	1	3.4	2	7.7	.48	.5	n.s.

b. Demographic and clinical characteristics of responders

	StudA (n = 60)		StudB (n = 60)				
	M	SD	*M*	SD	*t*	*p*	
Age (in years)	22.8	2.9	22.7	2.7	.2	.8	n.s.
Education (in years)	12	.4	11.9	.5	.59	.6	n.s.
BSI GSI	.6	.7	.7	.6	-.4	.7	n.s.
BSL-23 Score	.6	.7	.7	.7	-.7	.5	n.s.
	n	%	n	%	c^2	*p*	
Currently employed	40	67	46	77	1.48	.2	n.s.
In a relationship	30	50	32	53	.13	.7	n.s.
Children	0	0	0	0			n.s.

For group comparison between individuals with borderline personality disorder (BPD) and healthy controls (HCs) and between students paired with individuals with BPD (StudA) and students paired with HCs (StudB), *t*-tests and $\chi 2$-tests with level of significant of $p < .05$ were conducted for demographic and clinical characteristics. BSI=brief symptom inventory; BSL-23=borderline symptom list; GSI=global severity index; n.s.=not significant. * = significant at $p < .05$ ** = significant at $p < .01$.

Hypothesis 1: Frequency of coalition decisions

First, we checked whether patients with BPD preferred dyads over triads in comparison to healthy controls. About 30% of participants chose dyads over triads. Opposed to our hypothesis, there was no difference between Patients with BPD (in total 30.8%) and healthy controls (in total 28.7%) in either round (all p's > 0.05). See *Table 2a* for details.

Hypothesis 2: Bargaining behavior in dividing coalition values

Next, we assessed whether patients with BPD and healthy controls offered their co-players fair splits of the endowment. On average, participants offered about 47% of the dyad value and about 30% of the triad value two each responder. See Table 2b for details.

Behavior in dyads: In dyads, there were no differences in mean offer, rejection rate, and mean proposer earnings within rounds and between patients with BPD and healthy controls (all *p*'s > 0.05). All offers but one were accepted when dyads were formed.

Behavior in triads: The bargaining behavior in dividing coalition values in triads is shown in Fig. 2. While patients with BPD and healthy controls did not differ in mean offer, rejection rate, and mean proposer earnings during rounds 1 to 2 (all *p*'s > 0.05), healthy controls offered significantly lower amounts (HC 27.7% vs. BPD 32.1%, $t_{30.20} = -2.28$, $p = .03$) in round 3 (the final round) than patients with BPD.

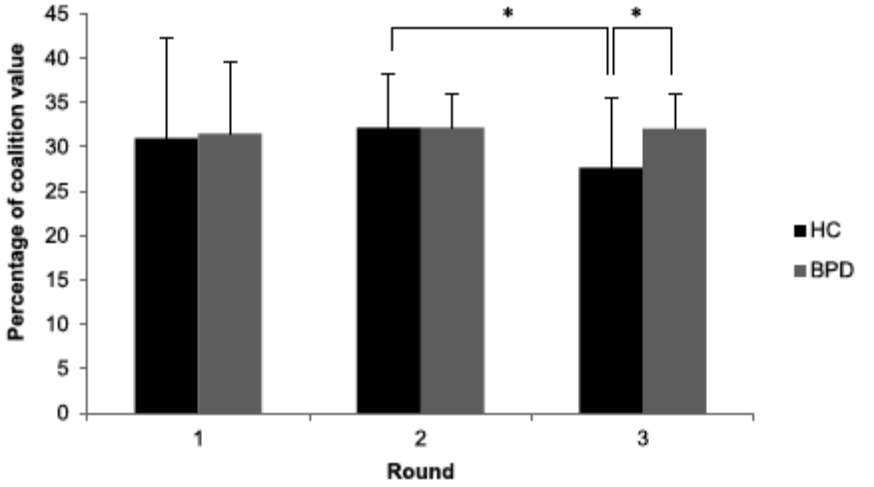


Figure 3. **Mean relative offer in triads.** Comparison of offers to responders in percentage of endowment (coalition value: 3000 points) between individuals with borderline personality disorder (BPD, *n*=26) and healthy controls (HCs, *n* = 29) in three rounds. Both responders received the same amount of points. Error bars indicate standard deviation. * = significant at $p < .05$, ** = significant at $p < .01$.

Hypothesis 3: Judgment and fairness ratings

Judgment: We verified whether the groups expected the recurrence of a partner strategy similarly and found a non-significant group by strategy by round interaction indicating that, in accordance with our hypothesis, the two groups did not differ in judgment. There were statistically significant main effects of strategy ($F_{2,89.36} = 80.93$, $p < .01$, $\eta^2 = 3175.775$) and round ($F_{7,268.64} = 3.98$, $p = .01$, $\eta^2 = 547.244$). Over all groups, the participants rated the likelihood of a triad to be chosen to be 30.8% by the exclusive strategy, 75.2% by the inclusive strategy, and 51% by the mixed strategy. There was a statistically significant two-way interaction of strategy and round ($F_{14,456.98} = 15.56$, $p = .01$, $\eta^2 = 753.74$). While participants

rated a decreasing likelihood of a triad to be chosen by the exclusive strategy (from 52.5% in round 1–20.9% in round 8), they rated an increasing likelihood for the inclusive strategy (from 58.3% in round 1–87.4% in round 8), and a fluctuating likelihood for the mixed strategy (around 50% during all rounds). Fig. 3 illustrates how participants anticipated the corresponding proposer type to select triads.

Fairness ratings: With regard to fairness ratings, the group by strategy interaction was also not statistically significant. Overall, the main effect of strategy was statistically significant ($F_{2,106} = 39.72$, $p < .01$, $\eta^2 = 1.27$). On a scale from 0 (= not at all) to 4 (= very much), patients with BPD and healthy controls rated inclusive strategies ($M = 3.4$, $SD = 0.1$) to be fairer than mixed strategies ($M = 2.4$, $SD = 0.1$), and mixed strategies to be fairer than exclusive strategies ($M = 1.4$, $SD = 0.2$, all p's < 0.01).

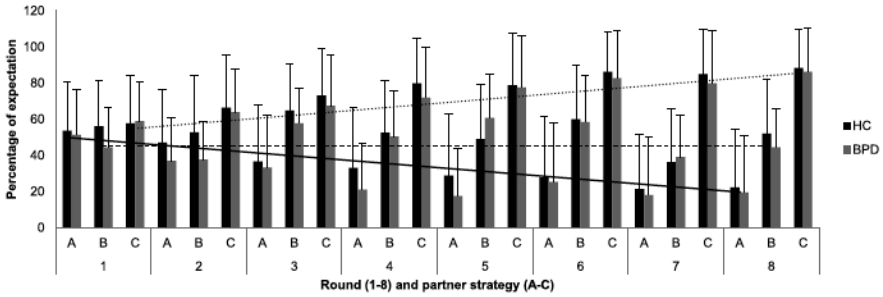


Figure 4. Partner strategy expectation for triads [%]. Comparison of partner strategy expectation for triads between individuals with borderline personality disorder (BPD, n=26) and healthy controls (HCs, n=29) as a function of partner strategies (A=choosing dyads in each round, B=randomly choosing either dyads or triads, C=choosing triads in each round) in eight rounds. Error bars indicate standard deviation. There was a statistically significant two-way interaction of strategy and round (p=.01). While the likelihood of the partner strategy expectation for a triad decreased during rounds while the partner displayed the exclusive strategy
(solid line), it increased while the partner displayed the inclusive strategy (dotted line), and fluctuated around 50% while the partner displayed the mixed strategy (dashed line).

Table 2. Coalition decisions

a. Frequency of coalition decicions						
		HC (*n* = 29)		BPD (*n* = 26)		
		n	%	*n*	%	c^2	*p*
Dyads in round 1		6	20.7	7	26.9	.30	.59
Dyads in round 2		11	37.9	9	34.6	.07	.8
Dyads in round 3		8	27.6	8	30.8	.07	.8
Overall dyads		25	28.7	24	30.8	.08	.8

b. Bargaining behavior in dividing coalition values							
		HC (*n* = 29)		BPD (*n* = 26)			
		M offer	SD	*M* offer	SD	*t*	*p*
Round 1	Dyads	46.1	8.6	57.1	18.9	-1.31	.216
	Triads	31.0	11.2	31.5	8.2	-0.18	.0857
Round 2	Dyads	46.1	10.8	47.4	4.4	-0.35	.730
	Triads	32.2	6.1	32.2	3.9	-0.01	.996
Round 3	Dyads	40.2	13.6	46.2	5.3	-1.16	.267
	Triads	27.7	7.9	32.1	3.9	-2.18	.036
		Rejections	%	Rejections	%	c^2	*p*
Round 1	Dyads	0	.0	0	.0	-	-
	Triads	4	17.4	2	10.5	.4	.53
Round 2	Dyads	0	.0	0	.0	-	-
	Triads	1	5.6	2	11.8	.43	.51
Round 3	Dyads	0	.0	1	12.5	1.07	.30
	Triads	4	21.1	0	.0	4.25	.04
		M win	SD	*M* win	SD	*t*	*p*
Round 1	Dyads	53.9	8.6	42.9	18.9	1.31	.22
	Triads	25.5	16.6	29.0	13.6	-.75	.46
Round 2	Dyads	53.9	10.8	52.6	4.4	.35	.73
	Triads	32.0	12.4	30.6	13.4	.32	.75
Round 3	Dyads	59.8	13.6	47.0	19.5	1.56	.14
	Triads	31.0	19.7	35.8	7.8	-.94	.35

For group comparison between individuals with borderline personality disorder (BPD) and healthy controls (HCs) *t*- and χ2-tests with level of significant of $p < .05$ were conducted for coalition decisions.

Hypothesis 4: Partner preference

There was a statistically significant difference in partner preference between patients with BPD and healthy controls ($c^2_{2,n=55} = 8.55$, $p < .05$). *Table 3* presents details. While the majority of healthy controls preferred partners with an inclusive strategy (triads; 86.2%), only half of patients with BPD (53.8%) opted for partners with an inclusive strategy. Whereas the minority of healthy controls (3.4%) selected

partners pursuing a mixed strategy, one third of patients with BPD (30.8%) did. Partners with an exclusive strategy (dyads) were chosen by one tenth of healthy controls (10.3%) and patients with BPD (15.4%).

Table 3. Preference for partner strategies

	HCs	BPD	Total
Dyads	3	4	7
Mixed	1	8	9
Triads	25	14	39
Total	29	26	55

BPD=borderline personality disorder; HCs=healthy controls.

DISCUSSION

In this study, we wanted to examine whether patients with BPD act fair in cases where no reciprocity concerns are relevant. Against our expectations, patients with BPD chose inclusion as often as healthy controls and, thereby, formed economically efficient and fair coalitions, i.e., they chose triads more often than dyads and offered their responders amounts of monetary units (MUs) similar to the amount offered by healthy controls. The observation that perception and execution of fairness seem to be equal in patients with BPD and healthy controls was also made when patients with BPD a) predicted and b) rated their partner's (un)fairness as did healthy controls. Notably, despite their equal judgment, nearly half of the patients with BPD would rather choose an unfair interaction partner (i.e., someone who pursuits either the two-person or mixed strategy) whereas the majority of healthy controls would choose a fair interaction partner. In other words, patients with BPD tend to rush headlong into disaster with their eyes wide open.

Unfavorable choices were also made in an Ultimatum game in which BPD responders accepted low offers at significantly higher rates than healthy responders[135]. This could be, in concordance with the *homo economicus*[e], interpreted as a strategy for maximal monetary gain on the one hand[128] and submissive acceptance of any (mal)treatment on the other hand. Matching the latter interpretation are observations that patients with BPD reject fair and generous offers of others[137,139]. The unfavorable partner preference of patients with BPD might be grounded in their negative and instable self-image due to childhood experiences of abuse of neglect[146]. These often traumatic childhood experiences and the frustration of basic childhood needs (e.g., secure attachment)

[e] "Homo economicus" is used in economics as a benchmark model of theoretical ideal of the consistently rational and self-interested agent.

have been linked to impairments in mentalizing (the ability to understand their own and others' mental states) which impact relationships negatively[147]. In prior studies, mentalizing abilities in patients with BPD have been predominantly tested by questionnaires, stories, and emotional facial stimuli (for a review, see[31]). These behavioral studies, despite some inconsistencies, show that patients with BPD have lower mentalizing abilities in complex tasks. For instance, they had difficulties in recognizing the intentions of others in video clips[148]. Interestingly, our results suggest that patients with BPD prefer unfavorable partners even if and when they accurately perceive the intentions underlying others' behavior. Another result of negative childhood experiences is the interference with the normative developmental process of integrating disparate mental representation of the self, relationships to others, and the world[149], hence early maladaptive schemas (self-defeating emotional, cognitive, and behavioral patterns)[150]. A cluster of schemas and coping styles is called schema mode. In heathy individuals, schema modes are mild, flexible mind states whereas, in individuals with personality disorder, schema modes are severe, rigid mind states. Particularly patients with BPD are characterized by a disorder-specific schema mode model[150,151] of which the *punitive parent mode,* i.e., self-hatred, shame, self-devaluation, and self-punishment, would explain the patients' behavior of submissive acceptance of maltreatment in our experiment. Indeed, patients with BPD have been found to evaluate themselves even more negatively after negative feedback whereas healthy controls did not change their self-evaluation[54]. Similarly, disagreement with the opinion of others provoked shame in patients with BPD but not in healthy controls[152]. However, in order to prevent participants form hedging monetary income[145], our study design did not include payoff consequences in the second task and, thus, does not allow straight-forward conclusions whether patients with BPD accept unfair offers. Thus, the transfer from our findings of partner preference to actual unfavorable partner choice in BPD is preliminary.

In the original set-up, the coalition formation game has been developed in order to study inefficiency and social exclusion in multilateral bargaining[140]. About one third of responders were excluded from bargaining and earned nothing. Similarly, we found only 30 percent of the BPD and healthy proposers to opt for dyads in our experiment. In accordance, it has been shown previously that patients with BPD engaged as much as HCs in altruistic punishment[134] and, exceedingly, showed solidarity with their unlucky co-players[138] which might be indicators of other-orientated empathic concerns when fairness is violated. In line with this, individuals with clinically relevant BPD features showed higher victim sensitivity than individuals without clinically relevant BPD features[153].

Our study employed three one-shot encounters in which each game round can be considered a first and only interaction with an unfamiliar partner. In one-shot encounters, strategic behavior with the intention of reciprocity should not matter[132]. In contrast, multiple encounters require trust and reciprocity which

patients with BPD seem to lack[128]. In one-shot encounters, there were no differences in bargaining behaviors between patients with BPD, patients with major depressive disorder, and healthy controls in trust and punishment games[154]. In our multi-round encounter study, patients with BPD behaved even fairer than healthy controls in the end round because, unlike healthy controls, they did not reduce their offer. The behavior of the healthy proposers was in line with the behavior of healthy volunteers in previous economic research studies where the so-called "free riding" is often observed to increase towards the end in finitely repeated games resulting in a large drop of the average level of investment in the final round[155]. In view of literature on the sensitivity of patients with BPD to unfairness, we have to see this persistent fair behavior as manifestation of social norms which patients with BPD follow in an inflexible manner which has been described before as "fierce determination for justice to prevail in all circumstances"[156].

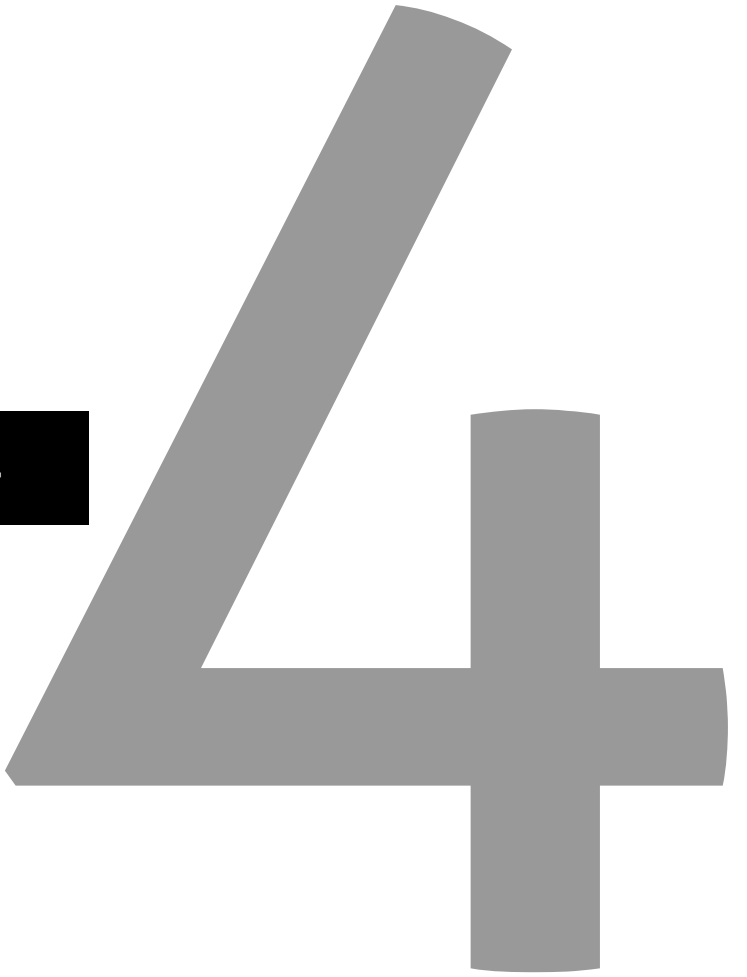
STRENGTH AND LIMITATIONS

This is the first study to assess equal and efficient coalition formation as well as judgment, fairness perception, and partner preference in patients with BPD. To increase the ecological validity of our procedure, participants engaged with human interaction partners. Nevertheless, we would like to address several limitations of our study. Since there are growing concerns regarding the replication of laboratory findings in clinical psychological research[122], our sample sizes of 26 patients with BPD and 29 healthy controls participants might have been too small. As mentioned in the introduction, three preliminary studies[137–139] with similar sample sizes have been carried out before which fit into the framework of this study. However, further replications with the current and similar paradigms are needed before strong conclusions can be drawn. Secondly, as many studies on BPD, we only tested female participants the majority of whom – typical of a naturalistic sample – patients with BPD had psychiatric comorbidities and took psychotropic medication[115]. Comparing the scores of the BSL-23 for disorder-specific symptoms with those of the validation study[143], we found comparable scores. Further, we did not include a patient control group so that we cannot draw conclusions specific for BPD. However, we matched our healthy controls on demographic variables. Additionally, our patients with BPD showed significantly higher values in the average total score (Global Severity Index) on the BSI which reflected a moderate level of psychiatric symptomatology and was similar to scores obtained in other studies of BPD[157,158]. Finally, the interaction partners of the proposers were all female students who did not undergo clinical assessment. Still, this should not be a problem since they are not in our focus of interest.

CONCLUSIONS

Our study provides preliminary evidence that the judgment of patients with BPD of what they deserve from others seems to be altered whereas the perception and execution of fairness are unaffected. This mirrors the previously described unfavorable partner choice of patients with BPD presumably due to a deeply negative self-image and low self-esteem. While clinical descriptions focus on dysfunctional relationships of patients with BPD with their significant others, recent economic-exchange games point to impairments in interaction also with unfamiliar partners. In the long term, research based on the current findings may help to stop the circle of negative self-image and low self-esteem that appears to underlie particularly rigid moral behavior to their own disadvantage. For instance, one interventional approach could be fostering self-compassion[159] as part of therapeutic interventions.

CHAPTER 4



Oxytocin improves facial emotion recognition in young adults with antisocial personality disorder

Marion Timmermann, Haang Jeung, Ruth Schmitt, Sabrina Boll, Christine M. Freitag, Katja Bertsch, Sabine C. Herpertz

Psychoneuroendocrinology, 2017

ABSTRACT

Deficient facial emotion recognition has been suggested to underlie aggression in individuals with antisocial personality disorder (ASPD). As the neuropeptide oxytocin (OT) has been shown to improve facial emotion recognition, it might also exert beneficial effects in individuals providing so much harm to the society. In a double-blind, randomized, placebo-controlled crossover trial, 22 individuals with ASPD and 29 healthy control (HC) subjects (matched for age, sex, intelligence, and education) were intranasally administered either OT (24 IU) or a placebo 45 min before participating in an emotion classification paradigm with fearful, angry, and happy faces. We assessed the number of correct classifications and reaction times as indicators of emotion recognition ability. Significant group × substance ×emotion interactions were found in correct classifications and reaction times. Compared to HC, individuals with ASPD showed deficits in recognizing fearful and happy faces; these group differences were no longer observable under OT. Additionally, reaction times for angry faces differed significantly between the ASPD and HC group in the placebo condition. This effect was mainly driven by longer reaction times in HC subjects after placebo administration compared to OT administration while individuals with ASPD revealed descriptively the contrary response pattern. Our data indicate an improvement of the recognition of fearful and happy facial expressions by OT in young adults with ASPD. Particularly the increased recognition of facial fear is of high importance since the correct perception of distress signals in others is thought to inhibit aggression. Beneficial effects of OT might be further mediated by improved recognition of facial happiness probably reflecting increased social reward responsiveness.

INTRODUCTION

Antisocial personality disorder (ASPD) is characterized by aggression, impulsivity, and reckless disregard for others' feelings[26]. Aggression and other maladaptive antisocial behaviors are commonly associated with deficits in social cognition in general and emotion recognition in particular[62,63,160–163]. For instance, individuals with ASPD show an impaired processing of faces indicated by deficient recognition of fearful and, to a lesser extent, sad and happy expressions[62,63] and a delayed response to fearful facial cues as compared to healthy controls (HC)[164]. Correctly recognizing emotional facial expressions is essential for successful everyday communication[165]. In particular, the eye region of faces conveys highly relevant cues for discrimination of emotions and therefore plays a crucial role in facial emotion recognition[166]. Eye gaze analysis found that individuals with antisocial traits tend to show reduced attentional shifts toward the eyes, which may be related to their deficits in fear recognition[167].

The neuropeptide oxytocin (OT) is a key mediator of complex emotional and social behaviors, including attachment, social recognition, and aggression. Intranasal OT administration has been shown to improve several aspects of social communication by increasing attention toward social cues in keeping with the social salience hypothesis[168–170]. Regarding emotion recognition, OT improves the perception of others' feelings[171] and enhances recognition of emotions in both static and dynamic facial expressions[172–177], which is essential for complex social interactions. The neuropeptide increases initial reflexive fixation changes toward the eyes[178]and appears to shift attention toward positive facial expressions[179]. However, several studies did not replicate improved facial emotion recognition[180,181] or modulated fixation changes after OT administration[175,181].
According to the social salience hypothesis, OT initiates affiliative behavior and modulates social perception by modifying the salience of various social cues suggesting that inconsistencies in previous data may result from variety in contextual aspects and interindividual differences[168,170]. Yet, other authors urge general clarification regarding the effectiveness of intranasal OT application[182], an administration form that is widely spread in human OT studies. Similarly, recent reports advised to critically view intranasal OT studies due to statistical and methodological limitations[183,184]. There is growing interest in understanding the role of OT in modulating social functioning in psychiatric disorders that are characterized by social impairments[185], such as ASPD. Thus, the aim of the current study was to investigate the effects of OT on facial emotion recognition in individuals with ASPD compared to HC in a double-blind, placebo-controlled crossover study. Using an emotion classification paradigm, we measured the proportion of correct answers and reaction times in response to fearful, angry, and happy facial expressions. We expected individuals with ASPD to perform worse than HC in recognizing fearful and to a lesser degree happy facial expressions[62],

and this poor performance to be compensated by the intranasal application of OT. In addition, we were interested in the OT effects on the processing of angry faces in young adults with ASPD as impaired angry face processing has been discussed to contribute to aggressive behavior[186] while OT reduced hypersensitivity to angry facial expressions in a clinical population[101]. Finally, we wanted to explore whether individuals with ASPD classify emotional faces better if the faces are initially presented on the eye region compared to the mouth region at the location of a former fixation cross as individuals with antisocial traits showed a reduced bias to shift attention toward the eyes[167].

METHODS AND MATERIALS

Participants

The present study included 22 mixed-sex young adults, diagnosed with ASPD as defined by DSM-IV criteria respectively DSM-5 criteria, and 29 HC subjects aged 18–30 years. Initially, 25 individuals with ASPD and 33 HC subjects were enrolled in the study. Four participants dropped out of the study not completing the second measurements (one ASPD and three HC). Two individuals with ASPD and one HC subject had to be excluded from further analyses because of technical problems (failure of behavioral data recording).

The ASPD group consisted of 8 women and 14 men with a mean age of 22.4 years (*SD* = 2.8). The HC group consisted of 11 women and 18 men with a mean age of 24.2 years (*SD* =4.1), matched with the ASPD group for age, sex, intelligence (IQ), and education. ASPD individuals were recruited at the Department of General Psychiatry at the University Hospital of Heidelberg, through local probation services, and through institutions offering anti-aggression trainings. The HC group was recruited through advertisement (black boards, online platforms, website of University Hospital of Heidelberg).

Exclusion criteria for all participants were an IQ < 80, pregnancy, breast feeding, claustrophobia, insufficient German language skills, any current medication (except oral contraceptives and levothyroxine), past cranio-cerebral injuries, and somatic, endocrine, or neurological diseases. For the ASPD group, further exclusion criteria were autism spectrum disorder, lifetime diagnoses of schizophrenia or bipolar disorder, and current alcohol or drug dependence. For the HC group, further exclusion criteria were any lifetime or current psychiatric disease, any psychological/psychiatric treatment, and antisocial behavior assessed by means of the Structured Clinical Interview for DSM-IV (SCID-I) and the International Personality Disorder Examination (IPDE).

The study was approved by the local ethics committee of the Faculty of Medicine, Heidelberg University, and conducted in accordance with the

Declaration of Helsinki. All participants gave written informed consent and were financially remunerated for their participation.

Measures

Diagnoses of ASPD as well as comorbidities with cluster B personality disorders (borderline and narcissistic personality disorders) were assessed by the IPDE, comorbidities with axis I disorders by the SCID-I. Among the most frequent comorbidities were substance use disorder (*N* = 16), alcohol use disorder (*N* = 10), previous mood disorder (*N* = 7), anxiety disorder (*N* = 6), current mood disorder (*N* = 4), posttraumatic stress disorder (N = 3), eating disorder (N =2), and somatic symptom disorder (*N* =1). Six females with ASPD had comorbid borderline personality disorder (BPD). Besides, 11 individuals (4 women, 7 men) out of the 22 individuals with ASPD met criteria for psychopathy according to the Self-Report Psychopathy Scale (SRP)[187], based on the recommended SRP cut-off score of 202 for diagnosis of psychopathy (Mokros, A., Hollerbach, P., Nitschke, J., and Habermeyer, E., unpublished observations). Four subtests of the Wechsler Adult Intelligence Scale were used to estimate intelligence. The State-Trait Anger Expression Inventory (STAXI)[188], Life History of Aggression (LHA)[189], and Reactive-Proactive Aggression Questionnaire (RPQ)[190] were questionnaires used to estimate the extent of aggressive behavior.

Experimental protocol

The study was conducted with a double-blind, placebo-controlled crossover design. Experiments took place at the University Hospital of Heidelberg and were part of a larger study. Participants were instructed to avoid food, drinks (except water), and nicotine two hours before starting the experiment. Initially, blood samples were taken for estradiol and progesterone assays analyzed by the Central Laboratory of the University Hospital of Heidelberg, and urine drug tests as well as pregnancy tests were conducted. All tested drugs required negative results except from tetrahydrocannabinol (THC) in individuals with ASPD due to the high coincidence of ASPD and THC usage (13 ASPD in the first measurement and 11 ASPD in the second measurement had positive THC results). Participants took part in a short training session to ensure they had understood the paradigm. OT (24 IU, Syntocinon Spray, Novartis, Basel, Switzerland) or placebo was intranasally administered with 6 puffs per nostril 45 min before starting the experiment as after this time period the effects of OT on the central nervous system are most effective[191,192]. The drug was prepared according to an externally computerized randomization list (simple randomization) by an independent pharmacist. At the first measurement occasion, 14 individuals with ASPD and 14 HC subjects received OT; the groups did not differ in substance order [$\chi 2(df=1, N = 51) = 1.19, p = 0.275$]. Before and after the experiment, participants were assessed for mood and alertness with a multidimensional mood questionnaire (MDBF)[193]. The two

measurements of each participant took place at a four-week interval. For women, regular four-week menstrual cycles (natural or pill-induced) were required for participation in order to conduct both measurements in the same cycle phase.

Emotion classification paradigm

The paradigm was conducted analogously to previous studies[101,178] and followed a 3× 2 design (facial emotions at full intensity: fearful, angry, happy; regions for initial face presentation: the faces were presented with either the eyes or the mouth at the location of a formerly presented fixation cross). Altogether, 216 emotional faces were presented in a randomized order in 3 blocks with 72 pictures each (36 female and 36 male faces). Pictures were selected from the NimStim Face Stimulus Set[194], the Pictures of Facial Affect[195], the Karolinska Directed Emotional Faces[196], and the Faces Database[197]. Faces were presented as black-and-white pictures with masked hair and ears.

The paradigm started with a fixation cross for two seconds, followed by the first facial expression with either the eyes or the mouth presented at the location of the former fixation cross. This task feature allowed differences between facial key regions to be observed. The face appeared for 150 ms and was followed by a black screen (1850 ms), and a fixation cross (2000–9000 ms). Participants were instructed to classify the facial emotion of each face as accurately and quickly as possible by pressing a corresponding button, indicating their emotion recognition ability. The measurements did not exceed 45 min, as levels of intranasally administered neuropeptides remain elevated within this period[191,198].

Data acquisition and analysis

Stimulus presentation and response recording was controlled with Presentation 14.2 (Neurobehavioral Systems, Albany, USA). Emotional faces were visible at 29 ° vertical and 65 ° horizontal visual angle. Regarding dependent variables, we determined the proportion of correct responses and reaction times for each condition (emotion and facial region). Proportion of correct responses was calculated by the number of correct answers divided by the number of all responded trials for each condition (mean response rate of individuals with ASPD: 96.9 percent; mean response rate of HC subjects: 99.2 percent). All participants achieved a minimum of 50 percent of correct responses in proportion to all presented trials for each emotion; therefore no data was excluded because of a high number of misclassifications. Reaction times were measured after stimulus onset for each condition. Only correctly classified trials were used for reaction time analyses as misclassifications could be attributed to impulsive reactions and therefore invalidate results. We controlled reaction times for outliers by removing all trials from further analysis that exceeded the individual mean reaction time plus or minus three standard deviations.

Statistical analysis

Statistical analyses were performed with IBM SPSS 22.0 (IBM, Armonk, USA) using repeated measures analyses of covariance (ANCOVAs) with the between-subject factor group (ASPD, HC), the within-subject factors substance (placebo, OT), facial emotion (fearful, angry, happy) and facial region (eyes, mouth), and the covariate age. The between-subject factor sex was included to control results for sex (due to the limited sample size, the effects of sex were not exploited). In an additional ANOVA, we compared results of ASPD individuals with THC usage to those of ASPD non-THC users. Normal distribution could be assumed due to Kolmogorov-Smirnov tests for all dependent variables in both groups except for the proportion of correct responses of fearful faces that were presented on the mouth in the OT condition. We used Huynh-Feldt correction in case of violations of the assumption of sphericity. For significant results, Dunn's multiple comparison procedure including Bonferroni correction for multiple testing was conducted post hoc. All statistical analyses were performed two-tailed with a level of significance of $p < 0.05$. Additionally, we applied non-parametric Mann-Whitney-tests for group comparisons and Wilcoxon-tests for within-group comparisons with Bonferroni-Holms correction for multiple testing.

RESULTS

Demographic and clinical characteristics

Table 4 provides detailed information about participants' demographic and clinical characteristics. After exclusion of $N = 7$ participants (3 individuals with ASPD with a mean age of 23.3 years (SD = 4.2) and a mean IQ of 103.0 (SD =10.6); 4 HC subjects with a mean age of 21.5 years (SD = 3.0) and a mean IQ of 103.3 (SD = 24.7)), the ASPD group tended to be younger than the HC group (age: $p = 0.058$). Thus, analyses of facial emotion classification were conducted with age as control factor. Individuals with ASPD and HC did not differ in estradiol and progesterone levels (measured only in women; *Table 4*). All subscales of the applied questionnaires indicated significantly higher scores for individuals with ASPD concerning behavior and aggression (Table 4).

Table 4. Demographic and clinical characteristics of individuals with ASPD and HC subjects

		Individuals with ASPD (n = 22)	HC subjects (n = 29)		
Demographic characteristics				F-/t-/χ^2-test[i,ii,iii]	p
Age (years)		22.36 (2.82)	24.24 (4.07)	–1.95	0.058
IQ (HAWIE-R)		98.41 (15.41)	105.66 (16.95)	–1.57	0.122
Male to female ratio[iii]		14:8	18:11	0.01	0.909
Clinical characteristics					
IPDE Antisocial		10.64 (3.76)	0.03 (0.19)	13.21	<0.001
STAXI	State anger	20.14 (7.27)	11.90 (2.50)	5.10	<0.001
	Trait anger	23.73 (5.28)	14.93 (3.52)	7.13	<0.001
	Anger IN	19.82 (4.26)	14.38 (3.48)	5.02	<0.001
	Anger OUT	20.32 (4.50)	12.07 (3.60)	7.27	<0.001
	Anger Control	18.59 (4.74)	24.38 (4.46)	–4.47	<0.001
LHA	Aggression	19.91 (4.42)	5.83 (4.29)	11.46	<0.001
	Antisocial behavior	10.27 (3.15)	1.59 (2.23)	11.54	<0.001
	Self-aggression	2.86 (2.44)	0.07 (0.26)	5.36	<0.001
	Total score	33.05 (6.22)	7.48 (5.81)	15.09	<0.001
	Reactive aggression	15.73 (3.15)	6.14 (3.38)	10.33	<0.001
	Proactive aggression	11.00 (5.44)	1.41 (1.24)	8.10	<0.001
	Interpersonal manipulation	54.27 (8.22)	36.17 (6.93)	8.52	<0.001
	Callous affect	45.73 (7.30)	35.79 (6.70)	5.05	<0.001
	Erratic lifestyle	55.45 (6.96)	39.66 (8.29)	7.21	<0.001
	Antisocial behavior	55.45 (6.96)	24.24 (7.18)	7.45	<0.001
	Total score	198.55 (22.00)	135.86 (20.75)	10.41	<0.001
Endocrinological data[ii] (only in women)					
Estradiol levesl (pg/ml)[iv]		45.14 (14.35)	42.30 (33.42)	0.10	0.752
Progestesterone levels (pg/ml)		1.28 (1.49)	0.99 (1.57)	0.32	0.577

Explanatory table legend: Data refers to means with standard deviation (SD) in brackets.
HAWIE-R = Wechsler Adult Intelligence Scale – Revised, German version (used for estimation of IQ); IPDE =International Personality Disorder Examination; STAXI =State-Trait Anger Expression Inventory; LHA = Life History of Aggression; RPQ = Reactive and Proactive Aggression Questionnaire; SRP =Self-Report Psychopathy Scale.
[i] For group comparisons between individuals with antisocial personality disorder (ASPD) and healthy control (HC) subjects, t-tests with level of significance of $p < 0.05$ were conducted for demographic and clinical characteristics, [ii] Univariate ANOVAs (F-tests) with a threshold set at $p < 0.05$ were conducted for endocrinological data, [iii] A χ.2-test was conducted to compare the male to female ratio of the ASPD group to the ratio of the HC group, [iv] Three estradiol values were below the minimum value of 11.8 pg/ml as determined by the laboratory.

Proportion of correct responses

All participants performed the emotion classification task with a high proportion of correct responses (Table 5). In addition to a trend for an overall group effect [$F(1,46) = 2.98$, $p = 0.091$, $\eta 2 = 0.06$] a significant three-way interaction of group, substance, and facial emotion [$F(2,92) = 4.59$, $p = 0.013$, $\eta 2 = 0.09$] was found (Fig. 1). In the placebo condition, individuals with ASPD performed significantly less accurately than HC in recognizing fearful ($p < 0.01$) and happy ($p < 0.05$) faces. However, in the OT condition no group differences for recognition of fear and happiness were found, indicating improved facial emotion recognition in the ASPD group. Non-parametric tests confirmed these results (group comparison of fearful faces presented on the eyes and mouth region: $Z = -1.96$, $p = 0.049$ and $Z = -2.14$, $p = 0.033$ in the placebo condition; $Z = -0.38$, $p = 0.703$ and $Z = -1.62$, $p = 0.105$ in the OT condition). In addition, the direct comparison between the placebo and the OT condition in the ASPD group revealed a significant increase of correctly classified fearful ($p < 0.01$) and happy faces ($p < 0.05$) following OT administration. Across groups, the three-way interaction yielded better classification of facial happiness than fear or anger in both the placebo and OT condition (all ps < 0.05). However, regarding the facial region for initial presentation, an interaction of group, emotion and facial region did not reach significance [$F(2,86) = 1.64$, $p = 0.201$, $\eta 2 = 0.04$]. The results of ASPD individuals with THC usage did not show any significant or trend-level differences compared to the results of ASPD non-THC users (all ps > 0.10).

Table 5. Proportion of correct responses and reaction times

		Individuals with ASPD (n = 22)		HC subjects (n = 29)	
		Placebo	Oxytocin	Placebo	Oxytocin
Proportion of correct responses					
Fearful	Eyes	0.817 (0.026)	0.892 (0.023)	0.905 (0.022)	0.909 (0.019)
	Mouth	0.792 (0.027)	0.848 (0.024)	0.897 (0.023)	0.987 (0.020)
Angry	Eyes	0.877 (0.024)	0.868 (0.030)	0.872 (0.020)	0.891 (0.025)
	Mouth	0.877 (0.027)	0.881 (0.024)	0.879 (0.023)	0.910 (0.020)
happy	Eyes	0.923 (0.027)	0.979 (0.009)	0.977 (0.023)	0.984 (0.008)
	Mouth	0.935 (0.027)	0.983 (0.011)	0.982 (0.023)	0.975 (0.009)
Reaction times (ms)					
Fearful	Eyes	1127 (51)	1111 (50)	1125 (44)	1067 (42)
	Mouth	1170 (53)	1132 (52)	1153 (45)	1102 (44)
Angry	Eyes	1114 (56)	1185 (71)	1176 (47)	1107 (60)
	Mouth	1113 (58)	1149 (62)	1193 (49)	1111 (53)
happy	Eyes	836 (30)	837 (32)	805 (26)	764 (27)
	Mouth	834 (31)	799 (32)	785 (26)	768 (27)

Explanatory table legend: Proportion of correct responses (upper part) and reaction times (in ms, lower part), subdivided into group (individuals with antisocial personality disorder (ASPD), healthy control (HC) subjects), substance (placebo, oxytocin), facial emotion (fearful, angry, happy), and facial region (eyes, mouth). Data refers to means with standard error of the mean (SEM) in brackets.

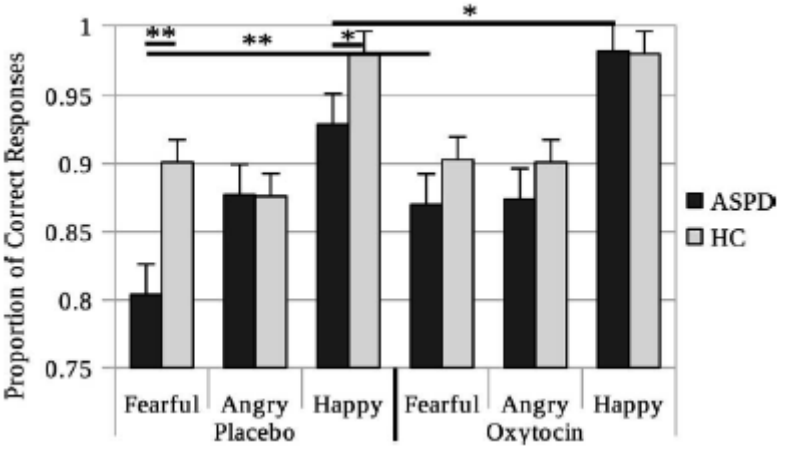


Figure 5. Proportion of Correct Responses as a Function of Facial Expression. Comparison of proportion of correct classifications between individuals with antisocial personality disorder (ASPD, N =22) and healthy control subjects (HC, N = 29) as a function of facial expression (fearful, angry, happy) in the placebo and oxytocin condition. Data were pooled across facial regions. Error bars indicate standard error of the mean.
*= significant at p 0.05, **= significant at p < 0.01.

Reaction times

Data yielded a significant interaction of group, substance, and facial emotion [$F(2,87) = 4.01$, $p = 0.023$, $\eta 2 = 0.08$]: In the placebo condition, the reaction times for angry faces differed significantly between the ASPD and HC group ($p < 0.05$) with HC subjects responding more slowly than individuals with ASPD (Fig. 2). This effect was mainly driven by the HC group with shorter reaction times for angry faces after OT administration compared to placebo ($p < 0.05$). Non-parametric tests confirmed these results (comparison of placebo and OT condition of HC subjects and angry faces presented on the eye region: $Z = -2.45$, $p = 0.014$; presented on the mouth region: $Z = -2.65$, $p = 0.008$). Descriptively, also longer reaction times in the ASPD group in the OT compared to the placebo condition contributed to the group differences in reaction times for angry faces without reaching significance (Table 5). Independently from the emotion, an interaction of group, substance, and facial region [$F(1,46) = 4.27$, $p = 0.044$, $\eta 2 = 0.09$] revealed longer reaction times in the ASPD group respectively shorter reaction times in the HC group (initial presentation on the eyes: $p < 0.01$; initial presentation on the mouth: $p < 0.05$) in the OT condition. Furthermore, HC subjects responded faster after OT administration compared to placebo for both on the eyes and on the mouth presented faces (ps < 0.01).

Across groups, participants showed shorter reaction times for happy faces than for fearful ($p < 0.01$) and angry ($p < 0.01$) faces [$F(2,90) = 3.64$, $p = 0.031$, $\eta 2 = 0.07$]. Similarly to the proportion of correct responses, the interaction of group ×emotion × facial region did not reach significance, $F(2,81) = 0.53$, $p = 0.568$, $\eta 2 = 0.01$.

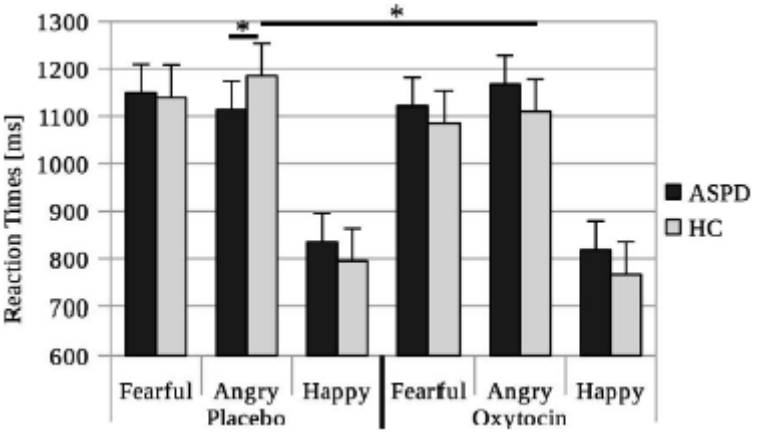


Figure 6. **Reaction Times as a Function of Facial Expression.** Comparison of reaction times between individuals with antisocial personality disorder (ASPD, N =22) and healthy control subjects (HC, N =29) as a function of facial expression (fearful, angry, happy) in the placebo and

oxytocin condition. Data were pooled across facial regions. Error bars indicate standard error of the mean.
* =significant at $p < 0.05$, ** =significant at $p < 0.01$.

DISCUSSION

This is the first study to investigate the effects of OT on facial emotion recognition in individuals with ASPD. In line with our expectations, the results revealed deficits in individuals with ASPD compared to HC in recognizing facial fear and happiness. Most interestingly, OT improved the classification of fearful faces in the ASPD group which no longer differed significantly from the HC group after OT administration. In addition, individuals with ASPD showed deficits in accurately classifying happy faces, which could also be compensated through administration of OT. In the HC group, OT administration was found to result in significantly shorter reaction times to angry faces, while the ASPD group tended to respond more slowly in the OT condition although the difference was not significant.

Proper recognition and processing of facial expressions are essential basics for social interaction[165]. The violence inhibition mechanism model[186] suggests that correct perception of distress signals such as fearful facial expressions in others is a key requirement for socialization and development of moral understanding[199]. Therefore, not being able to properly recognize emotions could contribute to the failure of antisocial individuals to inhibit aggression and other maladaptive behaviors. More specifically, according to the violence inhibition mechanism model, the underlying mechanism is thought to be the incapacity to form associations between emotional unconditioned stimuli (distress cues) and conditioned stimuli (representations of the acts, such as transgressions and rule breaking that caused the distress cues). The well replicated finding of amygdala dysfunction in antisocial individuals supports this theory[200], as well as recent behavioral data that point to deficits in the categorization of fearful expressions and its association with aggression in contrast to a general impaired visual perception[163]. This conclusion is also supported by findings that reported impairments in processing of fearful facial affects in various antisocial populations, such as offenders with autism spectrum disorders[201], patients with narcissistic disorders[202], and particularly youth with psychopathic traits (for review: [203]). Correspondingly, a meta-analysis of 20 studies reported a correlation between antisocial behavior and deficits in fear recognition[62], which was shown to be associated with an underlying amygdala dysfunction[200].

Interestingly, individuals with ASPD did not classify emotional faces better if the faces were presented with the eye region at the location of a formerly presented fixation cross as we had expected before due to eye gaze analyses which revealed a reduced attentional shift toward the eyes and might be related particularly to deficits in fear recognition[167]. As the amygdala plays an important

role in shifting attention toward the eyes[204], future studies may address the correlation of amygdala activation and the correct classification of facial emotions in antisocial populations with respect to the initially presented facial cue. Besides fear, deficits in processing of happy faces were observed in individuals with ASPD who classified happy faces less correctly than HC in our study. Notably, other authors, referring to psychopathic samples, described more general deficits in facial emotion recognition including happy faces[63]. Most interestingly, diminished husbands' sensitivity to fear and happiness in their wives' facial expressions were found to mediate the relationship between psychopathy and intimate partner violence[205]. The mechanisms which underlie deficits of facial recognition of happiness may lead to a lack of social reward responsiveness in antisocial individuals, implying a reduced sensitivity and less positive reactions to reward[206,207]. However, other authors could not replicate deficits in the recognition of happy faces in antisocial patients with and without psychopathic traits[160], calling for further studies which focus on the link between social reward responsiveness and the recognition of happy faces.

Interestingly, OT improved the classification of fearful and happy faces in individuals with ASPD. The improvement in classifying emotions by OT is consistent with a meta-analysis of 7 studies reporting an improvement of facial emotion recognition for fearful and happy faces after OT administration in healthy subjects[171] and a reduction of deficits in autism spectrum disorder[173,208]. The OT effects on facial emotion recognition may base on a modulated amygdala activity[173] with sex-specific effects having been reported[209]. In our study, improvement of facial emotion recognition was restricted to the ASPD group while no effect was found in the HC group. This finding is consistent with more recent findings of the effects of intranasal OT administration that point to the significance of interindividual differences in OT response[210] and gave rise to the theory that OT may compensate for trait-based deviations from adaptive social behavior rather than generally enhancing prosocial behavior[102,170]. Yet, with HC subjects classifying fearful and happy faces highly accurately in the placebo condition, the lack of an improvement of emotion recognition in HC subjects after OT administration may also relate to ceiling effects. In summary, OT improved recognition of fearful and happy faces in ASPD, however, due to ceiling effects we cannot evaluate whether OT also improved facial emotion recognition in HC subjects. Using similar emotion classification paradigms, studies on healthy participants as well as patients with BPD revealed similarly high proportions of correct responses compared to our HC subjects[101,178,204]. Thus, future studies may additionally implement facial expressions at lower intensities in order to avoid ceiling effects.

Our results may suggest clinically utilizable effects of OT. However, sharply in contrast with the hypothesis of a compensation for deviations from adaptive social behavior, intranasal OT administration also led to enhanced aggressive behavior in healthy participants[211]. Similarly, in a preliminary report with

6 individuals with ASPD, reactive aggression was partly increased after OT administration[68]. This study shows several limitations, such as a small number of participants and no placebo nor a comparison to a HC group. The contrary results may be traced back to a variety in contextual aspects and interindividual differences that influence the effects of OT[168,170].

In the OT condition, HC subjects yielded shorter reaction times in response to angry faces. OT administration has been reported to modulate reaction times of healthy volunteers in emotion classification tasks by other groups although the direction is inconsistent with shorter[176,181] and longer[172] reaction times. Contrary to our HC group, individuals with ASPD tended to respond more slowly to angry faces after OT administration. Although preliminary as the longer reaction times were not significant, they might indicate attenuating effects of OT administration in response to angry faces. Interestingly, our group found a reduction of amygdala reactivity in response to angry faces and normalization of originally faster initial fixation changes to the eyes of angry faces in BPD subsequent to intranasal OT administration, suggesting attenuating OT effects on a hyperreactivity to angry faces[101]. The comorbid BPD in 6 females with ASPD may have influenced our results regarding attenuating OT effects. Future studies are wanted that aim to elucidate shared and specific mechanisms of aggression between the two disorders[212]. This clarification should include a focus on the modulation of brain circuits, e.g. amygdala, anterior insula, and superior temporal gyrus, which play a crucial role in adaptive social behavior and empathy and have been hypothesized to be targets of OT effects (for review: [213]).

LIMITATIONS

While this study is the first to investigate OT effects on facial emotion recognition in individuals with ASPD, our study has several limitations. Besides the variation in demographic characteristics, the statistical power was not adequate ($1\text{-}\beta < 0.80$) to detect small differences between the groups due to the small sample size. As the small sample size furthermore led to mixed-sex groups, no sex-specific influences were investigated, but should be subject to further studies, the more as potential sex-specific differences in the effects of OT emphasize the importance of a direct comparison between women and men[209,214]. Moreover, ASPD is a very heterogeneous disorder that often shows comorbidities with other psychiatric disorders. With a large proportion of individuals with ASPD having comorbid BPD or psychopathic personality traits and the sample size being too small for comparisons between subgroups, our results cannot be attributed in a straightforward way to impairments specifically caused by ASPD. Similarly, the highly frequent THC usage in individuals with ASPD, which was also found in our sample, limits our results as animal studies revealed interactions between OT and the

endocannabinoid system[215] although we showed that THC usage did not contribute to the results of individuals with ASPD. Furthermore, as our study design was restricted to two negative emotions, it cannot contribute to the current research question whether deficits in facial emotion recognition are specific for fear or affect also other negative emotions[62,63,160,162], such as sadness or disgust. The implementation of emotional faces at full intensity may have led to highly accurate emotion classification and potential ceiling effects, particularly in healthy controls. Finally, the study design cannot differentiate between impairments in emotion recognition and emotion classification raising the question for future study designs which are suitable to detect the basic psychological processes that underlie impairments of facial processing in ASPD.

CONCLUSIONS

With deficits in facial emotion recognition presenting one of the underlying processes in antisocial behavior[216], our findings that OT ameliorated the recognition of fearful and happy emotions in young adults with ASPD may have clinical impact. Following the violence inhibition mechanism model, OT may exert anti-aggressive effects in young antisocial adults by improving the recognition of distress cues in potential victims and enhancing the reward value of happy facial cues. However, negative effects of OT administration such as increased aggressive behavior and increased out-group aggression also have been reported[68,211,217]. Before exploring OT as an anti-aggressive drug, our results have to be replicated not only in larger samples but also for long-term application. Several studies reported no further positive outcome (e.g. social interactions skills, emotion recognition ability, repetitive behaviors) in patients with autism spectrum disorder[174,218] and even negative side effects in mice[219] after long term OT administration, whereas recent studies also found positive OT effects in autistic children and patients with schizophrenia in the long run[220,221].

CHAPTER 5

5

Sex, dark traits, and leadership emergence

Haang Jeung-Maarse, Leo Altmeyer, Martin Vollmann, Simon Kirsch, Koen Schruers, Christiane Schwieren

Under Review

CHAPTER 6

6

General discussion

Parts of the general discussion have been extracted and adapted from reviews by Jeung-Maarse & Herpertz, 2020, Jeung et al., 2016, and Jeung et al., 2014.

GENERAL DISCUSSION

The previous chapters reported four studies in which behavioral experiments were developed, modified, or tested in individuals with personality disorders (PDs) or dark personality traits in order to study patterns of social interaction.

Chatroom paradigm (chapter 2)
In sociopsychology, the effects of social exclusion have been primarily studied in so-called Cyberball paradigm[43] in which participants toss a virtual ball to each other. Since findings are difficult to translate into real-life interactions, chatroom paradigms have been developed and tested in the last decade[111]. Instead of examining the effects of "liking" and "disliking" in classical Chatroom paradigms, we developed and tested a new, close to real-life interaction paradigm which confronted participants with peer feedback on personal opinions.

Economic-exchange games (chapter 3, 5)
In chapters 3 and 5, we modified experiments derived from behavioral economics and used real human interaction partners instead of fictitious partners (computer programs). Starting from basic computations of social exchange[248], game theory has fostered our understanding of the evolutionary and individual origins of cooperation and the assumption that repeated encounters shape cooperative and non-cooperative behavioral patterns in humans, evolutionary game theory found reciprocal strategies to be more successful than purely selfish strategies when computer strategies competed in a tournament, i.e., the "tit-for-tat" strategy[248] as both the simplest and most effective strategy for the Iterated Prisoner's Dilemma Game[249g]. In everyday life, however, there are not always repeated interactions between the same two individuals, but rather asymmetrical and indirect encounters between several members of a population. For instance, one person is in a position to help another but might not ask a favor in return, but other people might have observed their interaction. Therefore, helping others might improve reputation, and as a consequence, a helpful individual is more likely to receive help from (unrelated) others. Thus, indirect reciprocity promotes the evolution of

[g] The Iterated Prisoner's Dilemma Game is a simultaneous move game which is repeatedly played by two players. In one game round, each player can choose to cooperate or to defect. If both players cooperate, they both receive a reward for cooperating. If both players defect, they both receive a punishment. If one player defects while the other cooperates, the defector receives a higher reward for his/her defecting thanf or both of them cooperating. Similarly, the cooperator receives a higher punishment for his/her cooperating than for both of them defecting. Using the tit-for-tat strategy, a player cooperates at the first encounter and subsequently replicates the interaction partner's previous action. Ultimately, the strategy's success is based on the combination of being clear, nice, retaliatory, and forgiving if the other person makes a new attempt to cooperate[291]. Hence, direct reciprocity – as depicted by the tit-for-tat strategy– allowed the evolution of cooperation[292].

cooperation[250]. Since some individuals might be tempted to defect on others in favor of their own self-interest and exploit the altruism of others, the maintenance of cooperation typically requires a threat of punishment by a third party to keep the number of "free-riders" low[251,252]. As punishment could be costly for those who punish, this phenomenon is called altruistic punishment. For this purpose, in most human societies, legislature and executive and judicial powers enforce social norms that are agreed upon.

Facial emotion classification task (chapter 4)
In chapter 4, we examined the largest study group of individuals with ASPD in the literature. In order to study deficits in facial emotion recognition and oxytocin effects, we used a paradigm that had been already tested in healthy[178,204] and clinical[101] subjects. We specifically chose for this emotion classification task since it allows the investigation of early, reflexive processing of facial emotion expressions controlling for the initial fixation on eyes or mouth. The behavioral effects of oxytocin on emotion recognition have shown to be moderated by the way the faces were presented[171]: Types of faces (static[204,253], dynamic[254], morphed[255]) as well as exposure time (early[178] vs. late phase recognition[253]) of stimuli varied across neuroimaging studies. We did not use "neutral" faces as control as we wanted to avoid a hostile interpretation bias, which is the tendency to interpret ambiguous stimuli in a hostile manner, in individuals with ASPD[256]. However, the implementation of emotional faces at full intensity may have led to highly accurate emotion classification and potential ceiling effects, in both ASPD and HC groups.

FROM STUDIES TO THERAPIES

Current status of therapy

So far, there are eight European guidelines for the treatment of personality disorders (PDs), whereby these differ significantly in the diagnostic, psychotherapeutic and pharmacotherapeutic recommendations[275]. Five of the guidelines deal exclusively with BPD, one with ASPD and three with PD generally. Empirical evidence is only available for the treatment of BPD. In 2018, the guidelines of the National Institute for Health and Clinical Excellence (NICE) from 2009[276] were reviewed in a "surveillance report". Particularly detailed and well-founded guidelines on BPD are from Australia in 2013, which are explicitly based on the British NICE guidelines[277]. There is no evidence for the effectiveness of any pharmacotherapeutic treatment of BPD. The clinical reality is different: 85% of patients with BPD take at least one psychotropic drug, 20% of them at least four preparations[278].

All guidelines have in common that psychotherapy is the treatment is the first choice at BPD. Four elaborated therapeutic concepts of different theoretical orientation and duration of treatment have been proven to be effective[261,279]:

- Linehan's dialectical behavioral psychotherapy (DBT)[280]
- Bateman and Fonagy's mentalization-based therapy (MBT)[281]
- Young's schema therapy according to Young[150]
- Kernberg's transference-focused therapy (TFT)[282]

Overall, DBT had the most randomized controlled trials (RCTs), followed by MBT. While a moderate superiority of disorder-specific over non-disorder-specific psychotherapeutic interventions could be shown, there is insufficient evidence for the higher effectiveness of a disorder-specific program[283].

Future therapy strategies

A disorder-specific therapy does often not meet the requirements of the clinical reality in which individuals with PD face complex problems. In view of the longitudinal studies with criteriological remission but often persisting significant functional impairment, so-called dismantling studies try to identify effective individual components of more complex psychotherapy programs as independent functional units (modules)[284]. The so-called modular psychotherapy[285] builds on basic treatment skills and includes evidence-based unspecific techniques as well as specific techniques that target impairments in self and interpersonal functioning as well as prominent personality traits.

Future workplace strategies

As a practical application, organizations could account for dark personality traits in the context of personnel selection and internal promotions. For individual job performance, dark triad (DT) levels seem rather irrelevant[77], but for the organization as a whole and keeping in mind the well-being of colleagues and subordinates, identifying high-scoring individuals might be worth considering. There are also measuring instruments, which have been developed specifically for personality diagnostics in the context of personnel selection, such as the *Dark Triad Personality at Work*[286] to record DT characteristics, but these are again fraught with the problem of faking the desired impression. It should be taken into account that a healthy level of self-confidence or moderate DT expressions can be beneficial[287].

LIMITATIONS

Specific limitations of each study are reported in the discussion section of the respective chapters.

Sex matters

While we had mixed-sex samples in the studies examining individuals with ASPD and high DT traits (chapters 4-5), our findings in BPD are limited to female patients (chapters 2-3). Previously, it has been shown that sex matters in brain mechanisms underlying reactive aggression in BPD[95]. Therefore, follow-up studies should include male participants. In our ASPD study, we did not find differences between men and women. Yet, the sample size was too small to make a meaningful distinction. In the larger non-clinical student sample, we found evidence that socially agreeable behavior might influence socio-economic decision-making of women, especially those with low DT traits, but not men independent from their levels in DT traits.

Specificity

We used naturalistic study samples which meant the inclusion of co-morbidities and medication in the ASPD and BPD samples (chapters 2-4) and even no assessment of mental disorders in the student sample (chapter 5). We did not have a clinical control group in any of the studies, so effects might not be specific for the PDs and DT traits.

Another way to examine specific effects of a mental disorder could be a study sample with patients and healthy controls fulfilling categorical criteria as well as population-based individuals characterized in dimensional criteria. This approach could combine the best of two worlds of clinically well-described and relevant subjects with a limited number to recruit on the one hand, and a large sample of subjects who are subclinical and less well-described on the other hand.

Neurobiological correlates

All studies employed behavioral outcomes. Neurobiological correlates such as brain responses during the experiments have not been recorded. However, dealing with mental disorders means dealing with the complex organ that determines our personality and behavior: the brain. Thus, this thesis can be seen as preparatory work to establish experiments which can be performed in a brain scanner environment.

Summary
Samenvatting

SUMMARY

Personality disorders (PDs) have long been regarded as highly stable and treatment-resistant. However, PDs occur frequently and show high remission rates in the long term. Nevertheless, psychosocial recovery remains unsuccessful in a substantial proportion of cases. The latest revisions of the two prominent classification systems (DSM-5 and ICD-11) abandon the traditional view of specific personality disorders. Instead, the diagnostic process differentiates between three degrees of severity (mild, moderate, severe) and five prominent personality trait domains. Optionally, a borderline qualifying factor can be additionally codified. There is sufficient empirical evidence only for the treatment of borderline PD (BPD). Disorder-specific psychotherapy, in particular dialectic behavioral therapy (DBT) and mentalization-based therapy (MBT) have proven to be effective. Therapy modules targeting functional impairments and prominent personality trait domains could close the existing gaps in the disorder-specific treatment of PD. In particular, addressing interpersonal difficulties seems to be the core of all effective psychotherapy but also psychopharmacology such as the "love hormone" oxytocin.

With the revision of the classification systems, PDs are no longer viewed categorically but dimensionally. This means that personality traits can be found in a spectrum and fits with the literature that non-clinical impairments of interpersonal functioning have negative consequences for society, such as the adverse influence of dark personality traits on corporate cultures. It is unclear how this influence can be countered.

Both for the clinical and for societal implications, there is a need for the development of novel experimental paradigms which enable quantitative and standardized assessment of human social interaction in individuals with clinical personality disorders and sub-clinical dark personality traits. A general introduction to impairments in social interactions in individuals with personality disorders and dark personality traits is presented in **chapter 1**. This thesis compared behavioral outcomes of individuals with PDs/dark personality traits to individuals without PDs/dark personality traits in paradigms of social interaction, socio-economic exchange, and social cognition.

In **chapter 2**, we measured emotional responses to positive and negative compared to negative feedback in a chatroom. Individuals with BPD indicated more negative emotions, particularly shame, after negative feedback than healthy controls. Additionally, healthy controls showed a "happiness bias" while BPD patients reacted according to the valence of the feedback. When translating our findings to clinical practice, therapists should pay special attention to the activation and regulation of shame in the therapeutic process of patients with BPD.

In **chapter 3**, we found individuals with BPD to choose inclusion over exclusion up to the final round in a coalition formation game while healthy controls chose more exclusion in the end. Also different from the healthy controls, BPD

patients preferred partners with an exclusive strategy over a partner with an inclusive strategy. This finding, together with findings in chapter 1, pointed to a lower self-image as the basis for interactional difficulties in BPD which could be an interesting target for psychotherapeutical interventions other than emotion regulation such as fostering self-compassion.

Previously, hypersensitivity to social threat, i.e. angry faces, has been described and related to interpersonal difficulties in BPD. Moreover, oxytocin has been found to normalize abnormal amygdala reactivity to angry faces in BPD. In **chapter 4**, we extended these findings to individuals with antisocial personality disorder (ASPD) which is a distinct but related PD characterized by reactive aggression, and used a facial emotion classification taskunder placebo and. Similar to the findings in BPD, we found faster threat responses, i.e. shorter response latencies to angry faces, in individuals with ASPD compared to healthy controls. Moreover, we found deficits in the classification of fearful and happy faces in individuals with ASPD compared to healthy controls. Both behavioral differences could be abolished by oxytocin. Since both the attenuation of social threat sensitivity (angry faces), and correct perception of others' distress signals in others are thought to inhibit aggression, oxytocin might be beneficial in the treatment of ASPD.

In **chapter 5**, we further investigated the influence of non-clinical but so-called "dark" personality traits on socio-economic decision-making in dependence of sex and social setting in three economic-exchange games, namely the promotability game, the social value orientation task, and the delay-discounting task. High dark triad (DT) traits were associated with individualist behavior but not with rejection of tasks of low promotability nor delay discounting. In contrast, individuals low in DT traits exhibit mostly prosocial behavior. Interestingly, we found evidence that socially agreeable behavior might influence socio-economic decision-making of women, especially those with low DT traits, but not men. Ultimately, we did not ask about participants' self-image and societal role expectations, so we can only speculate about the reasons why women differ from men in mixed-sex settings but not single-sex settings.

In summary, this thesis contains novel insights into social interaction and provides experiments to study emotional and behavioral responses associated with impairments in interpersonal functioning. **Chapter 6** discusses the significance and implications of these findings and how they contribute to our understanding of personality pathology and assist in the development of new therapeutic approaches.

SAMENVATTING

Persoonlijkheidsstoornissen (PS's) worden al lang beschouwd als zeer stabiel en therapieresistent. PS's komen echter vaak voor en vertonen op de lange termijn hoge remissiepercentages. Toch blijft psychosociaal herstel in een substantieel deel van de gevallen onsuccesvol. De laatste herzieningen van de twee prominente classificatiesystemen (DSM-5 en ICD-11) verlaten de traditionele kijk op specifieke persoonlijkheidsstoornissen. In plaats daarvan maakt het diagnostisch proces onderscheid tussen drie graden van ernst (licht, matig, ernstig) en vijf prominente domeinen van persoonlijkheidskenmerken. Optioneel kan een borderline-kwalificerende factor aanvullend worden gecodificeerd. Alleen voor de behandeling van borderline PS (BPS) is voldoende empirisch bewijs. Stoornisspecifieke psychotherapie, in het bijzonder dialectische gedragstherapie (DBT) en mentaliserende therapie (MBT), zijn effectief gebleken. Therapiemodules gericht op functionele beperkingen en prominente domeinen van persoonlijkheidskenmerken zouden de bestaande hiaten in de stoornisspecifieke behandeling van PS's kunnen dichten. Met name het aanpakken van interpersoonlijke problemen lijkt de kern te zijn van alle effectieve psychotherapie, maar ook psychofarmacologie zoals het "liefdeshormoon" oxytocine.

Met de herziening van de classificatiesystemen worden PS's niet langer categorisch maar dimensionaal bekeken. Dit betekent dat persoonlijkheids-kenmerken in een spectrum voorkomen en dit past bij de literatuur dat niet-klinische stoornissen in het interpersoonlijk functioneren negatieve gevolgen hebben voor de samenleving, zoals de nadelige invloed van donkere persoonlijkheidskenmerken op bedrijfsculturen. Het is onduidelijk hoe deze invloed kan worden tegengegaan.

Zowel voor de klinische als voor de maatschappelijke implicaties is er behoefte aan de ontwikkeling van nieuwe experimentele paradigma's die kwantitatieve en gestandaardiseerde beoordeling van menselijke sociale interactie mogelijk maken bij personen met klinische persoonlijkheidsstoornissen en subklinische donkere persoonlijkheidskenmerken. Een algemene inleiding tot stoornissen in sociale interacties bij personen met persoonlijkheidsstoornissen en donkere persoonlijkheidskenmerken wordt gepresenteerd in hoofdstuk 1. Dit proefschrift vergeleek gedragsuitkomsten van individuen met PS's/donkere persoonlijkheidskenmerken met individuen zonder PS's/donkere persoonlijkheids-kenmerken in paradigma's van sociale interactie, sociaal-economische uitwisseling en sociale cognitie.

In hoofdstuk 2 hebben we emotionele reacties op positieve en negatieve feedback in een chatroom gemeten in vergelijking met negatieve feedback. Mensen met een BPS gaven meer negatieve emoties aan, met name schaamte, na negatieve feedback vergeleken met mensen in de gezonde controlegroep.

Bovendien vertoonden gezonde controles een "geluksbias", terwijl BPS-patiënten reageerden op basis van de valentie van de feedback. Bij het vertalen van onze bevindingen naar de klinische praktijk dienen therapeuten speciale aandacht te besteden aan het activeren en reguleren van schaamte in het therapeutische proces van patiënten met een borderline-stoornis.

In hoofdstuk 3 ontdekten we dat personen met een borderline-stoornis inclusie verkiezen boven uitsluiting tot aan de laatste ronde in een coalitievormingsspel, terwijl gezonde controles uiteindelijk meer uitsluiting kozen. Eveneens verschillend van de gezonde controles, gaven BPS-patiënten de voorkeur aan partners met een exclusieve strategie boven een partner met een inclusieve strategie. Deze bevinding, samen met de bevindingen in hoofdstuk 1, wees op een lager zelfbeeld als basis voor interactieproblemen bij BPS, wat een interessant doelwit zou kunnen zijn voor andere psychotherapeutische interventies dan emotieregulatie, zoals het bevorderen van zelfcompassie.

Voorheen was overgevoeligheid voor sociale dreiging, d.w.z. boze gezichten, beschreven en gerelateerd aan interpersoonlijke problemen bij BPS. Bovendien is gevonden dat oxytocine de abnormale amygdala-reactiviteit op boze gezichten bij BPS normaliseert. In hoofdstuk 4 breidden we deze bevindingen uit naar personen met antisociaale PS (ASPS), een aparte maar gerelateerde PS die wordt gekenmerkt door reactieve agressie. We gebruikten een classificatietaak voor gezichtsemoties (facial emotion classification task). Onder placebo vonden wij kortere reactielatenties op boze gezichten, bij personen met ASPS in vergelijking met gezonde controles. Bovendien vonden we tekortkomingen in de classificatie van angstige en blije gezichten bij personen met een ASPS in vergelijking met gezonde controles. Beide gedragsverschillen zouden door oxytocine kunnen worden opgeheven. Aangezien wordt aangenomen dat zowel de verzwakking van de gevoeligheid voor sociale dreiging (boze gezichten) als de juiste perceptie van de noodsignalen van anderen bij anderen agressie remmen, kan oxytocine gunstig zijn bij de behandeling van ASPS.

In hoofdstuk 5 onderzochten we verder de invloed van niet-klinische maar zogenaamde “donkere” persoonlijkheidskenmerken op sociaal-economische besluitvorming in afhankelijkheid van geslacht en sociale omgeving in drie economische uitwisselingsspellen, namelijk het promotabiliteitsspel (promotability game), de sociale-waarde-oriëntatietaak (social value orientation task) en de vertragingsdisconteringstaak (delay discounting task). Hoge donkere triade (DT) eigenschappen waren geassocieerd met individualistisch gedrag, maar niet met het afwijzen van taken met een lage promotabiliteit of het uitstellen van korting. Daarentegen vertonen individuen met lage DT-kenmerken overwegend prosociaal gedrag. Interessant is dat we bewijs vonden dat sociaal aangenaam gedrag de sociaal-economische besluitvorming van vrouwen zou kunnen beïnvloeden, vooral die met lage DT-kenmerken, maar niet van mannen.

Uiteindelijk hebben we niet gevraagd naar het zelfbeeld en de sociale rolverwachtingen van de deelnemers, dus we kunnen alleen speculeren over de redenen waarom vrouwen verschillen van mannen in omgevingen van gemengd geslacht, maar niet in omgevingen van hetzelfde geslacht.

Samenvattend bevat dit proefschrift nieuwe inzichten in sociale interactie en biedt het experimenten om emotionele en gedragsreacties te bestuderen die verband houden met stoornissen in het interpersoonlijk functioneren. Hoofdstuk 6 bespreekt de betekenis en implicaties van deze bevindingen en hoe ze bijdragen aan ons begrip van persoonlijkheidspathologie en helpen bij de ontwikkeling van nieuwe therapeutische benaderingen.

Contributions and impact
Bijdragen en impact

CONTRIBUTIONS AND IMPACT

Contributions

Addressing interpersonal difficulties is the best way to manage individuals with PDs and dark personality traits. In ASPD and BPD, previous studies have mainly studied the meaning of anger on social interactions and deficits in cognitive control including emotion regulation and inhibitory control[288]. However, low interpersonal functioning cannot be sufficiently explained by affective dysregulation and impulsivity, but it has also been attributed to impaired social cognition, such as impaired facial emotion recognition[29,62]. These impairments in social cognition and resulting pattern of unstable relationships in adulthood could be linked to adverse childhood experiences that alter the neuropeptide systems such as the oxytocin system[289]. Consistent with this, early social environment-induced changes in the oxytocin system were shown to be associated with the social phenotype implying that social functioning might be under epigenetic control probably via deoxyribonucleic acid (DNA) methylation of the oxytocin receptor[290].

Yet, studies that present emotions in faces only show a part of what constitutes interpersonal interactions. Therefore, there is a need for the application of experimental paradigms which enable quantitative and standardized assessment of human social interaction. Since social interactions are not only the subject of studies in psychiatry but also in other disciplines such as psychology, sociology, and economics, we pursuit an interdisciplinary approach in this thesis in order to characterize various aspects of social interaction influenced by personality pathology.

In particular, the contributions of this thesis are:

- the development of a chatroom paradigm which captures emotional responses that may arise in everyday interactions (chapter 2)
- the modification of the coalition formation task which examines partner preferences (chapter 3)
- the use of a well-established facial emotion classification task in the largest ASPD sample in the literature so far (chapter 4)
- the modification of the promotability game, social value orientation task, and delay discounting task in order to examine the influence of sex and dark personality traits on leadership emergence (chapter 5)

Impact paragraph

So far, the focus of research and therapy for BPD has been primarily on emotion regulation (DBT) and mentalization (MBT). This thesis contributes to the understanding that a lower self-image might contribute substantially to interpersonal problems up to problematic partner choices in BPD. Whether self-

compassion has a positive effect on interpersonal relationships would be an interesting research question for intervention studies.

For ASPD, there is even no evidence for effective therapy. Social threat hypersensitivity seems to be a common basis for reactive aggression in BPD and ASPD. This thesis shows that individuals with ASPD seem to be less susceptible to distress in others which is consistent with findings in psychopaths. Oxytocin seems to have a positive effect on both PDs. However, a comparison between the two PDs can only be made with reservations, since we used the same study design (an emotion classification task in a double-blind, randomized, placebo-controlled crossover trial in patients versus healthy controls) but not the same study sample as the prior study in BPD. Further, the study in BPD patients also examined amygdala reactivity which would be the next step in the examination of individuals with ASPD.

Finally, we derived the influence of personality traits on leadership emergence. However, a follow-up study should include actual leaders. It would be particularly interesting whether societal role expectations would dominate socio-economic decision-making more than personality traits and social value orientation. While economists rather study interaction behavior in large student or community samples, comprising real interactions between uninformed participants, clinicians compare the behaviors of two distinct groups in predefined and programmed settings. In this thesis, we bring the strengths of the two approaches together. This meant that we engaged (pre)clinical subjects in real interactions with non-clinical controls instead of fictitious partners. This is all the more interesting since the reaction of the interaction partners to the behavior of the index participants is also measured which could even give an indication of the countertransference therapists seem to struggle with. Therefore, future studies with blinded participation could address socioeconomic decision-making in patients and therapists equally.

BIJDRAGEN EN IMPACT

Bijdragen

Het aanpakken van interpersoonlijke problemen is de beste manier om om te gaan met personen met PS's en donkere persoonlijkheidskenmerken. Bij ASPS en BPS hebben eerdere studies voornamelijk de betekenis van woede op sociale interacties en tekorten in cognitieve controle, waaronder emotieregulatie en remmende controle, bestudeerd [288]. Een laag interpersoonlijk functioneren kan echter niet voldoende worden verklaard door affectieve disregulatie en impulsiviteit, maar wordt ook toegeschreven aan een verminderde sociale cognitie, zoals een verminderde herkenning van gezichtsemoties[29,62]. Deze stoornissen in sociale cognitie en het daaruit voortvloeiende patroon van onstabiele relaties op volwassen leeftijd kunnen in verband worden gebracht met ongunstige ervaringen in de kindertijd die de neuropeptidesystemen zoals het oxytocinesysteem veranderen[289]. In overeenstemming hiermee bleken vroege door de sociale omgeving veroorzaakte veranderingen in het oxytocinesysteem geassocieerd te zijn met het sociale fenotype, wat impliceert dat sociaal functioneren mogelijk onder epigenetische controle staat, waarschijnlijk via deoxyribonucleïnezuur-(DNA-)methylatie van de oxytocinereceptor[290].

Toch laten onderzoeken die emoties in gezichten weergeven slechts een deel zien van wat interpersoonlijke interacties zijn. Daarom is er behoefte aan de toepassing van experimentele paradigma's die kwantitatieve en gestandaardiseerde beoordeling van menselijke sociale interactie mogelijk maken. Aangezien sociale interacties niet alleen het onderwerp zijn van studies in de psychiatrie, maar ook in andere disciplines zoals psychologie, sociologie en economie, streven we in dit proefschrift naar een interdisciplinaire benadering om verschillende aspecten van sociale interactie beïnvloed door persoonlijkheids-pathologie te karakteriseren.

De bijdragen van dit proefschrift zijn in het bijzonder:

- de ontwikkeling van een chatroom-paradigma dat emotionele reacties vastlegt die kunnen ontstaan in alledaagse interacties (hoofdstuk 2)
- de aanpassing van de coalitievormingstaak die partnervoorkeuren onderzoekt (hoofdstuk 3)
- het gebruik van een gevestigde classificatietaak voor gezichtsemoties (facial emotion classification task) in de grootste ASP-steekproef tot nu toe in de literatuur (hoofdstuk 4)
- de modificatie van het promotabiliteitsspel (promotability game), de sociale-waarde-oriëntatietaak (social value orientation task) en de vertragingsdisconteringstaak (delay discounting task) om de invloed van geslacht en donkere persoonlijkheidskenmerken op het ontstaan van leiderschap te onderzoeken (hoofdstuk 5)

Impact paragraph

Tot nu toe lag de focus van onderzoek en therapie voor BPS vooral op emotieregulatie (DBT) en mentaliseren (MBT). Dit proefschrift draagt bij aan het inzicht dat een lager zelfbeeld substantieel kan bijdragen aan interpersoonlijke problemen tot problematische partnerkeuzes bij BPS. Of zelfcompassie een positief effect heeft op interpersoonlijke relaties zou een interessante onderzoeksvraag zijn voor interventiestudies.

Voor de ASPS is zelfs geen bewijs voor een effectieve therapie. Overgevoeligheid voor sociale dreiging lijkt een algemene basis te zijn voor reactieve agressie bij BPS en ASPS. Dit proefschrift laat zien dat personen met een ASPS minder vatbaar lijken te zijn voor angst bij anderen, wat consistent is met bevindingen bij psychopaten. Oxytocine lijkt een positief effect te hebben op beide PS's. Een vergelijking tussen de twee PS's kan echter alleen onder voorbehoud worden gemaakt, aangezien we hetzelfde onderzoeksontwerp gebruikten (een emotieclassificatieparadigma in een dubbelblinde, gerandomiseerde, placebogecontroleerde cross-over studie bij patiënten versus gezonde controles) maar niet dezelfde onderzoekssteekproef als de voorafgaande studie in BPS. Verder onderzocht de studie bij BPS-patiënten ook de amygdala-reactiviteit, wat de volgende stap zou zijn in het onderzoek van personen met ASPS.

Ten slotte hebben we de invloed afgeleid van persoonlijkheids-kenmerken op het ontstaan van leiderschap. Een vervolgonderzoek zou echter echte leiders moeten omvatten. Het zou met name interessant zijn of maatschappelijke rolverwachtingen de sociaal-economische besluitvorming meer zouden domineren dan persoonlijkheidskenmerken en sociale waardenoriëntatie. Terwijl economen liever interactiegedrag bestuderen in grote steekproeven van studenten of gemeenschappen, bestaande uit echte interacties tussen niet-geïnformeerde deelnemers, vergelijken clinici het gedrag van twee verschillende groepen in vooraf gedefinieerde en geprogrammeerde instellingen. In dit proefschrift brengen we de sterke punten van de twee benaderingen samen. Dit betekende dat we (pre)klinische proefpersonen in echte interacties betrokken met niet-klinische controles in plaats van met fictieve partners. Dit is des te interessanter omdat ook de reactie van de interactiepartners op het gedrag van de indexdeelnemers wordt gemeten, wat zelfs een indicatie zou kunnen geven van de tegenoverdracht waar therapeuten mee lijken te worstelen. Daarom zouden toekomstige studies met geblindeerde deelname de sociaaleconomische besluitvorming bij patiënten en therapeuten in gelijke mate kunnen aanpakken.

ACKNOWLEDGMENTS

My work is dedicated to all patients who continually ignite my wonder and passion about our innate need for social belonging. It is in loving memory of my godson **Benito**.

First of all, I would like to thank **Koen Schruers**, who not only championed my master's thesis but now serves as the guiding force behind this endeavor as well. Your infectious enthusiasm for education has not only nurtured me as a scientist but also as a psychotherapist. I also wish to acknowledge **Christiane Schwieren**, whose invaluable input and insights have enriched the depth of my scientific inquiry. Your exemplary leadership has left an indelible mark on my journey, and I am profoundly grateful to have crossed paths with you.

To my paranymphs, **Joost Maarse** and **Ines Ulrich**, you have been my steadfast companions through life's peaks and valleys. Your unwavering curiosity, openness to diverse perspectives, and relentless pursuit of knowledge inspire me to be a better version of myself each day. Thank you for teaching me the art of forging connections and drawing out the best within me.

My sincere appreciation goes to my supervisors and collaborators **Katja Bertsch** and **Sabine C. Herpertz** for their patience, steady guidance, and unending support. The colloborative efforts of my co-authors – **Lea Altmeyer**, **Sabrina Boll**, **Christine M. Freitag**, **Simon Kirsch**, **Christoph W. Korn**, **Falk Mancke**, **Mike M. Schmitgen**, **Ruth Schmitt**, **Marion Timmermann**, **Martin Vollmann**, and **Stephan Walther** – have been pivotal in collecting, analyzing, and interpreting the data. I extend my gratitude to **Konstantin Bülau**, **Klarissa Ehlers**, **Mareike Ochner**, **Arturo Olivares Rivera**, **Paula Posiege** and **Johanna Sander** for their invaluable contributions to patient recruitment and data management. A special note of appreciation to **Christina Baacke**, who kept the ASPD study running which was a particular task of its own.

I wish to express my deep thanks to **Michiel Bolkestein**, **Asha M. Das**, and **Massimiliano Grassi** whose excellent theses have served as guiding beacons. My mentors – **Martin Driessen**, **Manuela Dudeck**, **Frank Schneider**, and **Knut Schnell** – have significantly shaped my personal and professional growth. Further, I am indepted to the **Evangelisches Klinikum Bethel**, **Medical Faculty OWL** and the **University of Bielefeld** for providing me with a new academic home and supporting my journey through the Female Clinician Scientist Fellowship and mentorship program movement.med.

My research would have been inconceivable without the constant and loving support of my husband **Joost Maarse**, also my paranymph. His discerning insights as the first reader and reviewer of my manuscripts have been invaluable. My heartfelt appreciation goes to our children, **Esther Hye-Jin** and **Jakob San-Gyu**, for the joy, and boundless positivity they infuse into my life. I would not have been able to finish my thesis without the engagement and encouragement of my parents, **Hyeu-Kyung Lim** and **Han-Kyu Jeung**, and my sister **Bit-Na Fischer** with **Tim Kyu** and **Mia Lim**. I would also like to

thank my Dutch family for their support: **Tiny** and **Hans Maarse**, **Anke** and **Remco Tragter** with **Liz** and **Jim**, **Ireen Maarse** and **Erik Heringa**. I reserve a special place of gratitude for **Anke**, whose meticulous proofreading has been invaluable.

To the remarkable women – scientists, clinicians, and psychotherapists – whose friendships have been a wellspring of inspiration and strength: **Stephanie Behringer**, **Sheila Cordeiro**, **Curie Kure**, **Henrike Möller** with my godson **Nicklas**, **Elise Sarpaczki**, **Regine Scheib**, **Isabella Schneider**, **Nicole Seidner**, **Lara Wahlster**, and **Sarah Wahlster** with my goddaughter **Sophia**.

CURRICULUM VITAE

Personal data

Name: Haang Jeung-Maarse, née Jeung
Date of Birth: February, 21, 1982
Place of Birth: Herford, Germany

Education

09/10-07/14 Maastricht University
- International Master of Affective Neuroscience (M.Sc.)
- Thesis: "Social Emotions: Loneliness and Social Bonds in Borderline Personality Disorder"
- Supervisor: Prof. K. Schruers

05/11 University of Bonn
- Medical Doctor (MD) degree (Dr. med.)
- Dissertation: "Amelioration of the declarative memory by the NMDA receptor partial agonist D-cycloserine in healthy individuals"
- Supervisor: Prof. W. Maier

2008 Heidelberg University Medical School
- Final state exam (Staatsexamen)
- Clinical rotations in Heidelberg, Salamanca and Cambridge

2001 Ratsgymnasium Bielefeld
- High school diploma (Abitur)

Professional practice

07/22-present Department of Psychiatry, University of Bielefeld
- Attending physician (Oberärztin) and researcher

10/19–06/22 Department of Psychiatry, University of Saarland, Homburg
- Deputy director (Leitende Oberärztin)

02/11–09/19 Department of Psychiatry, University of Heidelberg
- Attending physician (Oberärztin) since 08/16
- Specialist physician (Fachärztin) since 03/15
- Resident (Assistenzärztin) of psychiatry

02/09–01/11 Department of Psychiatry, University of Bonn
- Resident of psychiatry
- Spokesperson for the residents of psychiatry from 08/09 until 07/10

Qualifications
- Management of psychiatric and psychosomatic hospitals, Mibeg Institute Cologne (03/22)
- Clinical supervisor of psychiatry and psychotherapy, German Association for Psychiatry, Psychotherapy, and Psychosomatics (DGPPN; 06/18)
- Leadership training across professions, University of Heidelberg (05/18)
- Baden-Württemberg Certificate for teaching and learning at university level (10/16)
- Congress organization, German Society for Biological Psychiatry (DGBP; 2012, 2018)
- Selection committee, German Academic Scholarship Foundation (since 2013)

Research institutions
The research described in this thesis was performed at University of Heidelberg (chapters 2-5), University of Bielefeld (chapter 5), and Maastricht University (chapter 5).

Awards
- Female Clinician Scientist Fellowship, University of Bielefeld, Medical Faculty (02/23-02/24)
- Research grant, Baden-Württemberg Foundation (2019)
- Rahel Goitein-Straus fellowship for young female scientists, University of Heidelberg, Medical Faculty (10/15-07/16)
- Scholarship of the Board for the International Master of Affective Neuroscience, Maastricht University (09/10)
- Scholarship, German National Academic Foundation (Studienstiftung des deutschen Volkes) 05/04-12/08

PUBLICATIONS

Jeung-Maarse, H., Altmeyer, L., Vollmann, M., Kirsch, S., Schruers, K., & Schwieren, C. (2023). Sex, dark traits, and leadership emergence, *under review*.

Jeung-Maarse, H., & Herpertz, S. C. (2020). Neues zur Diagnostik und Therapie von Persönlichkeitsstörungen – Änderungen in ICD-11. New insights into diagnostics and therapy of personality disorders—Changes in ICD-11. *Der Nervenarzt*. https://doi.org/10.1007/s00115-020-00936-7

Jeung, H., Vollmann, M., Herpertz, S. C., & Schwieren, C. (2018). Consider others better than yourself: Social decision-making and partner preference in Borderline Personality Disorder. *Journal of Behavior Therapy and Experimental Psychiatry*. https://doi.org/10.1016/j.jbtep.2018.11.004

Jeung, H., Walther, S., Korn, C. W., Bertsch, K., & Herpertz, S. C. (2018). Emotional responses to receiving peer feedback on opinions in borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment*, *9*(6). https://doi.org/10.1037/per0000292

Herpertz, S. C., Bertsch, K., & **Jeung, H.** (2018). Neurobiology of Criterion A: self and interpersonal personality functioning. *Current Opinion in Psychology*, *21*. https://doi.org/10.1016/j.copsyc.2017.08.032

Timmermann, M., **Jeung, H.**, Schmitt, R., Boll, S., Freitag, C. M., Bertsch, K., & Herpertz, S. C. (2017). Oxytocin improves facial emotion recognition in young adults with antisocial personality disorder. *Psychoneuroendocrinology*, *85*. https://doi.org/10.1016/j.psyneuen.2017.07.483

Hirjak, D., Thomann, A. K., Kubera, K. M., Wolf, R. C., **Jeung, H**., Maier-Hein, K. H., & Thomann, P. A. (2017). Cortical folding patterns are associated with impulsivity in healthy young adults. *Brain Imaging and Behavior*, *11*(6). https://doi.org/10.1007/s11682-016-9618-2

Jeung, H., Schwieren, C., & Herpertz, S. C. (2016). Rationality and self-interest as economic-exchange strategy in borderline personality disorder: Game theory, social preferences, and interpersonal behavior. *Neuroscience and Biobehavioral Reviews*, *71*. https://doi.org/10.1016/j.neubiorev.2016.10.030

Jeung, H., Thomann, P. A., & Wolf, R. C. (2015). Novel gene variations in early-onset frontotemporal dementia with positive family history of neural ceroid lipofuscinosis-1. *Neurology: Clinical Practice*, *5*(6). https://doi.org/10.1212/CPJ.0000000000000134

Jeung, H., & Herpertz, S. C. (2014). Impairments of interpersonal functioning: Empathy and intimacy in borderline personality disorder. *Psychopathology*, *47*(4). https://doi.org/10.1159/000357191

Herpertz, S. C., **Jeung, H.**, Mancke, F., & Bertsch, K. (2014). Social dysfunctioning and brain in borderline personality disorder. *Psychopathology*, *47*(6).

https://doi.org/10.1159/000365106

Otte, D.-M., Barcena de Arellano, M. L., Bilkei-Gorzo, A., Albayram, Ö., Imbeault, S., **Jeung, H.**, Alferink, J., & Zimmer, A. (2013). Effects of Chronic D-Serine Elevation on Animal Models of Depression and Anxiety-Related Behavior. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0067131

Onur, O. A., Schlaepfer, T. E., Kukolja, J., Bauer, A., **Jeung, H.**, Patin, A., Otte, D. M., Shah, N. J., Maier, W., Kendrick, K. M., Fink, G. R., & Hurlemann, R. (2010). The N-Methyl-D-Aspartate Receptor Co-agonist D-Cycloserine Facilitates Declarative Learning and Hippocampal Activity in Humans. *Biological Psychiatry*, *67*(12). https://doi.org/10.1016/j.biopsych.2010.01.022

BOOK CHAPTERS

Roesch-Ely, D., Rek, I., **Jeung-Maarse, H**., Bartolovic M (2018): Dabei hab' ich niemandem was getan. In: Jünger J: Ärztliche Kommunikation: Praxisbuch zum Masterplan Medizinstudium 2020. Schattauer, S. 312-8.

Bartolovic, M., **Jeung-Maarse, H.**, Roesch-Ely, D. (2018): Das Leben macht keinen Sinn mehr. In: Jünger J: Ärztliche Kommunikation: Praxisbuch zum Masterplan Medizinstudium 2020. Schattauer, S. 319-26.

TRANSLATIONS

First, M.B.: Handbuch der Differenzialdiagnosen - DSM-5®: Deutsche Ausgabe herausgegeben von Winfried Rief (2016). Hogrefe Verlag, S. 149-65 und S. 334-48. ISBN-10: 3801727572. ISBN-13: 978-3801727574.

American Psychiatric Association: Diagnostisches und Statistisches Manual Psychischer Störungen DSM-5®: Deutsche Ausgabe herausgegeben von Peter Falkai und Hans-Ulrich Wittchen (2015). Hogrefe Verlag, S. 883-940 und S. 1045-70. ISBN 978-3-8017-2599-0

REFERENCES

1. Pincus, A. L., Cain, N. M. & Halberstadt, A. L. Importance of Self and Other in Defining Personality Pathology. *Psychopathology* **53**, 133–140 (2020).
2. Borkenau, P. & Ostendorf, F. *NEO-Fünf-Faktoren Inventar:(NEO-FFI); nach Costa und McCrae*. (Hogrefe, 1993).
3. Samuels, J. Personality disorders: Epidemiology and public health issues. *Int. Rev. Psychiatry* **23**, 223–233 (2011).
4. Ellison, W. D., Rosenstein, L. K., Morgan, T. A. & Zimmerman, M. Community and Clinical Epidemiology of Borderline Personality Disorder. *Psychiatric Clinics of North America* vol. 41 561–573 (2018).
5. Grilo, C. M. *et al.* Two-year stability and change of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *J. Consult. Clin. Psychol.* **72**, 767–775 (2004).
6. Zanarini, M. C., Frankenburg, F. R., Reich, D. B. & Fitzmaurice, G. Time to attainment of recovery from borderline personality disorder and stability of recovery: A 10-year prospective follow-up study. *Am J Psychiatry* **167**, 663–667 (2010).
7. Gunderson, J. G. *et al.* Ten-year course of borderline personality disorder: psychopathology and function from the Collaborative Longitudinal Personality Disorders study. *Arch Gen Psychiatry* **68**, 827–837 (2011).
8. Cramer, V., Torgersen, S. & Kringlen, E. Personality disorders and quality of life. A population study. *Compr. Psychiatry* **47**, 178–184 (2006).
9. Tyrer, P., Reed, G. M. & Crawford, M. J. Personality disorder 1. Classification, assessment, prevalence, and effect of personality disorder. *Lancet* **385**, (2015).
10. Herpertz, S. C. [A new approach to classifying Personality Disorders]. *Fortschr. Neurol. Psychiatr.* **86**, 150–155 (2018).
11. Regier, D. A. *et al.* DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. *Am. J. Psychiatry* **170**, 59–70 (2013).
12. Widiger, T., Trull, T., Clarkin, J. & Sanderson, C. A description of the DSM-IV personality disorders with the five-factor model of personality. (2002).
13. Anderson, J., Snider, S., Sellbom, M., … R. K.-P. & 2014, undefined. A comparison of the DSM-5 Section II and Section III personality disorder structures. *Elsevier*.

14. Herpertz, S. C. Empathy and personality disorders from a neurobiological perspective. *Forensische Psychiatrie, Psychologie, Kriminologie* vol. 12 192–198 (2018).
15. Renneberg, Babette; Böttcher, J. Wenn Vermeidung das soziale Leben bestimmt. Diagnose und kognitive Verhaltenstherapie der Ängstlich-vermeidenden Persönlichkeitsstörung. *Psychotherapeutenjournal* **2**, 116–122 (2017).
16. Olajide, K. *et al.* Development and Psychometric Properties of the Standardized Assessment of Severity of Personality Disorder (SASPD). *J. Pers. Disord.* **32**, 44–56 (2018).
17. Zimmermann, J. *et al.* The structure and correlates of self-reported DSM-5 maladaptive personality traits: Findings from two german-speaking samples. *J. Pers. Disord.* **28**, 518–540 (2014).
18. Bach, B. & First, M. B. Application of the ICD-11 classification of personality disorders. *BMC Psychiatry* **18**, (2018).
19. Stern, A. Psychoanalytic investigation and therapy in the borderline group of neuroses. *Psychoanal Q* **7**, 467–489 (1938).
20. Grinker, R., Werble, B. & Drye, R. The borderline syndrome: A behavioral study of ego functions. *New York, Basic Books* (1968).
21. Gunderson, J. G. & Singer, M. T. Defining borderline patients: An overview. *Am J Psychiatry* **32**, 1–10 (1975).
22. Gunderson, J. G. & Kolb, J. E. Discriminating features of borderline patients. *Am J Psychiatry* **135**, 792–796 (1978).
23. Spitzer, R. L., Endicott, J. & Williams, J. B. Research diagnostic criteria. *Arch Gen Psychiatry* **36**, 1381–1383 (1979).
24. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Third Edition. *Washington, D. C., APA* (1980).
25. World Health Organisation. ICD-10 Classifications of Mental and Behavioural Disorder: Clinical Descriptions and Diagnostic Guidelines. *Geneva, WHO.* (1992).
26. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fith Edition. *Washington, D. C., APA.* (2013).
27. Sadikaj, G., Moskowitz, D. S., Russell, J. J., Zuroff, D. C. & Paris, J. Quarrelsome behavior in borderline personality disorder: influence of behavioral and affective reactivity to perceptions of others. *J Abnorm Psychol* **122**, 195–207 (2013).
28. Daros, A. R., Zakzanis, K. K. & Ruocco, A. C. Facial emotion recognition in borderline personality disorder. *Psychol Med* 1–11 (2012) doi:10.1017/S0033291712002607.
29. Domes, G., Schulze, L. & Herpertz, S. C. Emotion recognition in

borderline personality disorder-a review of the literature. *J Pers Disord* **23**, 6–19 (2009).
30. Stiglmayr, C. E. *et al.* Aversive tension in patients with borderline personality disorder: a computer-based controlled field study. *Acta Psychiatr Scand* **111**, 372–379 (2005).
31. Jeung, H. & Herpertz, S. C. Impairments of interpersonal functioning: Empathy and intimacy in borderline personality disorder. *Psychopathology* **47**, (2014).
32. Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M. & Bohus, M. Borderline personality disorder. *Lancet* **364**, 453–461 (2004).
33. Putnam, K. M. & Silk, K. R. Emotion dysregulation and the development of borderline personality disorder. *Dev Psychopathol* **17**, 899–925 (2005).
34. Carpenter RW, T. T. J. Components of emotion dysregulation in borderline personality disorder: a review. *Curr Psychiatry Rep* 335 (2013) doi:10.1007/s11920-012-0335-2.
35. Jovev, M. *et al.* Attentional processes and responding to affective faces in youth with borderline personality features. *Psychiatry Res* **199**, 44–50 (2012).
36. Minzenberg, M. J., Fisher-Irving, M., Poole, J. H. & Vinogradov, S. Reduced Self-Referential Source Memory Performance is Associated with Interpersonal Dysfunction in Borderline Personality Disorder. *J Pers Disord* **20**, 42–54 (2006).
37. Domes, G. *et al.* Recognition of facial affect in Borderline Personality Disorder. *J Pers Disord* **22**, 135–147 (2008).
38. Wagner, A. W. & Linehan, M. M. Facial expression recognition ability among women with borderline personality disorder: implications for emotion regulation? *J Pers Disord* **13**, 329–344 (1999).
39. Arntz, A. & Veen, G. Evaluations of others by borderline patients. *J Nerv Ment Dis* **189**, 513–521 (2001).
40. Nigg, J. T., Lohr, N. E., Western, D., Gold, L. J. & Silk, K. R. Malevolent object representations in borderline personality disorder and major depression. *J Abnorm Psychol* **101**, 61–67 (1992).
41. Lazarus, S. A., Cheavens, J. S., Festa, F. & Zachary Rosenthal, M. Interpersonal functioning in borderline personality disorder: a systematic review of behavioral and laboratory-based assessments. *Clin Psychol Rev* **34**, 193–205 (2014).
42. Lis, S. & Bohus, M. Social interaction in borderline personality disorder. *Curr Psychiatry Rep* **15**, 338 (2013).
43. Williams, K. D. & Jarvis, B. Cyberball: a program for use in research on

interpersonal ostracism and acceptance. *Behav Res Methods* **38**, 174–180 (2006).
44. De Panfilis, C., Riva, P., Preti, E., Cabrino, C. & Marchesi, C. When social inclusion is not enough: Implicit expectations of extreme inclusion in borderline personality disorder. *Pers. Disord* **6**, 301–309 (2015).
45. Domsalla, M. *et al.* Cerebral processing of social rejection in patients with borderline personality disorder. *Soc Cogn Affect Neurosci* **9**, 1789–1797 (2014).
46. Renneberg, B. *et al.* Perception of social participation in borderline personality disorder. *Clin Psychol Psychother* **19**, 473–480 (2012).
47. Staebler, K. *et al.* Facial emotional expression in reaction to social exclusion in borderline personality disorder. *Psychol Med* **41**, 1929–1938 (2011).
48. Lawrence, K. A., Allen, J. S. & Chanen, A. M. Impulsivity in borderline personality disorder: reward-based decision-making and its relationship to emotional distress. *J Pers Disord* **24**, 786–799 (2010).
49. Cherek, D. R., Moeller, F. G., Dougherty, D. M. & Rhoades, H. Studies of violent and nonviolent male parolees: II. Laboratory and psychometric measurements of impulsivity. *Biol Psychiatry* **41**, 523–529 (1997).
50. Dougherty, D. M., Bjork, J. M., Huckabee, H. C., Moeller, F. G. & Swann, A. C. Laboratory measures of aggression and impulsivity in women with borderline personality disorder. *Psychiatry Res* **85**, 315–326 (1999).
51. McCloskey, M. S. *et al.* Evaluation of behavioral impulsivity and aggression tasks as endophenotypes for borderline personality disorder. *J Psychiatr Res* **43**, 1036–1048 (2009).
52. New, A. S. *et al.* Laboratory induced aggression: a positron emission tomography study of aggressive individuals with borderline personality disorder. *Biol Psychiatry* **66**, 1107–1114 (2009).
53. Ruocco, A. C. *et al.* Medial prefrontal cortex hyperactivation during social exclusion in borderline personality disorder. *Psychiatry Res* **181**, 233–236 (2010).
54. Korn, C. W., La Rosée, L., Heekeren, H. R. & Roepke, S. Social feedback processing in borderline personality disorder. *Psychol Med* **46**, 575–587 (2016).
55. Dulz, B., Briken, P., Kernberg, O. F. & Rauchfleisch, U. Handbuch der Antisozialen Persönlichkeitsstörung. *Stuttgart: Schattauer* (2017).
56. Hare, R. D. Psychopathy. *Theory Res.* (1970).
57. Hare, R. D. The Hare psychopathy checklist-revised: Manual. *Multi-*

Health Syst. Inc. (1991).
58. Hare, R. D. Manual for the revised psychopathy checklist. (2003).
59. Black, D., Baumgard, C., American, S. B.-B. O. T. & 1995, U. The long-term outcome of antisocial personality disorder compared with depression, schizophrenia, and surgical conditions. *Bull Am Acad Psychiatry Law* **23**, 43–52 (1995).
60. Werner, K. B., Few, L. R. & Bucholz, K. K. Epidemiology, comorbidity, and behavioral genetics of antisocial personality disorder and psychopathy. *Psychiatr. Ann.* **45**, 195–199 (2015).
61. Young, M. H., Justice, J. V & Erdberg, P. *Assault in Prison and Assault in Prison Psychiatric Treatment*. *J Forensic Sci* vol. 49 www.astm.org (2004).
62. Marsh, A. A. & Blair, R. J. R. Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience and Biobehavioral Reviews* vol. 32 454–465 (2008).
63. Dawel, A., O'Kearney, R., McKone, E. & Palermo, R. Not just fear and sadness: meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neurosci. Biobehav. Rev.* **36**, 2288–304 (2012).
64. Blair, R. J. R. *et al.* Reduced sensitivity to others' fearful expressions in psychopathic individuals. *Pers. Individ. Dif.* **37**, 1111–1122 (2004).
65. Dargis, M., Wolf, R. C. & Koenigs, M. Psychopathic traits are associated with reduced fixations to the eye region of fearful faces. *J. Abnorm. Psychol.* **127**, 43–50 (2018).
66. Blair, R. J. R. The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience* vol. 14 786–799 (2013).
67. Gedeon, T., Parry, J. & Völlm, B. The role of oxytocin in antisocial personality disorders: A systematic review of the literature. *Frontiers in Psychiatry* vol. 10 (2019).
68. Alcorn, J. L., Rathnayaka, N., Swann, A. C., Moeller, F. G. & Lane, S. D. Effects of Intranasal Oxytocin on Aggressive Responding in Antisocial Personality Disorder. *Psychol. Rec.* **65**, 691–703 (2015).
69. Babiak, P., Neumann, C. S. & Hare, R. D. Corporate psychopathy: Talking the walk. *Behav Sci Law* **28**, 174–193 (2010).
70. Furtner, M. R., Maran, T. & Rauthmann, J. F. Dark Leadership: The Role of Leaders' Dark Triad Personality Traits. *Lead. Dev. Deconstructed* 75–99 (2017) doi:10.1007/978-3-319-64740-1_4.
71. Grijalva, E., Psychology, D. N.-A. & 2015, undefined. Narcissism and counterproductive work behavior (CWB): Meta-analysis and consideration of collectivist culture, Big Five personality, and

narcissism's facet structure. *Wiley Online Libr.* **64**, 93–126 (2014).
72. Jonason, P. K., Webster, G. D., Schmitt, D. P., Li, N. P. & Crysel, L. The Antihero in Popular Culture: A Life History Theory of the Dark Triad. *Rev. Gen. Psychol.* **16**, 192–199 (2012).
73. Paulhus, D. L. & Williams, K. M. The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *J. Res. Pers.* **36**, 556–563 (2002).
74. Christie, R. & Geis, F. L. *Studies in Machiavellianism*. (Academic Press, 1970).
75. Raskin, R. N. & Hall, C. S. A narcissistic personality inventory. *Psychol. Rep.* **45**, 590 (1979).
76. Hare, R. D. Comparison of Procedures for the Assessment of Psychopathy. *J. Consult. Clin. Psychol.* **53**, 7–16 (1985).
77. O'Boyle, E. H., Forsyth, D. R., Banks, G. C. & McDaniel, M. A. A meta-analysis of the Dark Triad and work behavior: A social exchange perspective. *J. Appl. Psychol.* **97**, 557–579 (2012).
78. Connelly, B. S., Lilienfeld, S. O. & Schmeelk, K. M. Information exchange article integrity tests and morality: Associations with ego development, moral reasoning, and psychopathic personality. *Int. J. Sel. Assess.* **14**, 82–86 (2006).
79. Landay, K., Harms, P. D. & Credé, M. Shall we serve the dark lords? A meta-analytic review of psychopathy and leadership. *J. Appl. Psychol.* **104**, 183–196 (2019).
80. Zettler, I. & Solga, M. Not Enough of a 'Dark' Trait? Linking Machiavellianism to Job Performance. *Eur. J. Pers.* **27**, 545–554 (2013).
81. Spain, S. M., Harms, P. D. & Wood, D. Stress, well-being, and the dark side of leadership. in *The role of leadership in occupational stress* (eds. Gentry, W. A. & Clerkin, C.) 33–60 (Emerald Group Publishing, 2016). doi:10.1108/S1479-355520160000014002.
82. Volmer, J., Koch, I. K. & Göritz, A. S. The bright and dark sides of leaders' dark triad traits: Effects on subordinates' career success and well-being. *Pers. Individ. Dif.* **101**, 413–418 (2016).
83. Weaver, S. G. & Yancey, G. B. the Impact of Leadership Styles on Organizational Commitment and Turnover. *Leadersh. Rev.* **10**, 104–124 (2010).
84. Başar, U. A Multilevel Study of Relationships between Leaders Dark Triad and Employee Burnout: Mediating Role of Perceived Dark Leadership. *J. Bus. Res. - Turk* **12**, 2407–2423 (2020).
85. Tucker, A. W. The Mathematics of Tucker: A Sampler. *Two-Year Coll. Math. J.* **14**, 228 (1983).

86. Deutchman, P. & Sullivan, J. The Dark Triad and framing effects predict selfish behavior in a one-shot Prisoner's Dilemma. *PLoS One* **13**, e0203891 (2018).
87. Mokros, A. *et al.* Diminished Cooperativeness of Psychopaths in a Prisoner's Dilemma Game Yields Higher Rewards. *J. Abnorm. Psychol.* **117**, 406–413 (2008).
88. Curry, O., Chesters, M. J. & Viding, E. The psychopath's dilemma: The effects of psychopathic personality traits in one-shot games. *Pers. Individ. Dif.* **50**, 804–809 (2011).
89. Rilling, J. K. *et al.* Neural Correlates of Social Cooperation and Non-Cooperation as a Function of Psychopathy. *Biol. Psychiatry* **61**, 1260–1271 (2007).
90. Malesza, M. The effects of the Dark Triad traits in prisoner's dilemma game. *Curr. Psychol.* **39**, 1055–1062 (2020).
91. Coid, J. W. Aetiological risk factors for personality disorders. *Br. J. Psychiatry* **174**, 530–538 (1999).
92. Herpertz, S. C., Bertsch, K. & Jeung, H. Neurobiology of Criterion A: self and interpersonal personality functioning. *Curr. Opin. Psychol.* **21**, (2018).
93. Hidalgo, N. A. I. *et al.* Time course of facial emotion processing in women with borderline personality disorder: an ERP study. *J. psychiatry Neurosci. JPN* **41**, 16 (2016).
94. Cullen, K. R. *et al.* Brain activation in response to overt and covert fear and happy faces in women with borderline personality disorder. *Brain Imaging Behav.* **10**, 319–331 (2016).
95. Herpertz, S. C. *et al.* Brain Mechanisms Underlying Reactive Aggression in Borderline Personality Disorder-Sex Matters. *Biol. Psychiatry* **82**, 257–266 (2017).
96. Brislin, S. J. *et al.* Callousness and affective face processing in adults: Behavioral and brain-potential indicators. *Personal. Disord.* **9**, 122–132 (2018).
97. Decety, J., Skelly, L., Yoder, K. J. & Kiehl, K. A. Neural processing of dynamic emotional facial expressions in psychopaths. *Soc. Neurosci.* **9**, 36–49 (2014).
98. Pera-Guardiola, V. *et al.* Brain Structural Correlates of Emotion Recognition in Psychopaths. *PLoS One* **11**, e0149807 (2016).
99. Kendrick, K. M., Guastella, A. J. & Becker, B. Overview of human oxytocin research. in *Current Topics in Behavioral Neurosciences* vol. 35 321–348 (Springer Verlag, 2018).
100. Yang, X., Wang, W., Wang, X. T. & Wang, Y. W. A meta-analysis of

hormone administration effects on cooperative behaviours: Oxytocin, vasopressin, and testosterone. *Neurosci. Biobehav. Rev.* **126**, 430–443 (2021).
101. Bertsch, K. *et al.* Oxytocin and reduction of social threat hypersensitivity in women with borderline personality disorder. *Am. J. Psychiatry* **170**, 1169–1177 (2013).
102. Bartz, J. *et al.* Oxytocin can hinder trust and cooperation in borderline personality disorder. *Soc Cogn Affect Neurosci* **6**, 556–563 (2011).
103. Jeung-Maarse, H. & Herpertz, S. C. Neues zur Diagnostik und Therapie von Persönlichkeitsstörungen – Änderungen in ICD-11New insights into diagnostics and therapy of personality disorders—Changes in ICD-11. *Nervenarzt* (2020) doi:10.1007/s00115-020-00936-7.
104. Gunderson, J. G. *et al.* Ten-year course of borderline personality disorder: psychopathology and function from the Collaborative Longitudinal Personality Disorders study. *Arch Gen Psychiatry* **68**, 827–837 (2011).
105. Beeney, J. E., Hallquist, M. N., Clifton, A. D., Lazarus, S. A. & Pilkonis, P. A. Social disadvantage and borderline personality disorder: A study of social networks. *Pers. Disord* (2016) doi:10.1037/per0000234.
106. Hepp, J. *et al.* Interpersonal problems and negative affect in Borderline Personality and Depressive Disorders in daily life. *Clin Psychol Sci* **5**, 470–484 (2017).
107. Scott, L. N. *et al.* Borderline personality disorder symptoms and aggression: A within-person process model. *J Abnorm Psychol* **126**, 429–440 (2017).
108. Dixon-Gordon, K. L., Gratz, K. L., Breetz, A. & Tull, M. A laboratory-based examination of responses to social rejection in borderline personality disorder: the mediating role of emotion dysregulation. *J Pers Disord* **27**, 157–171 (2013).
109. Downey, G. & Feldman, S. I. Implications of rejection sensitivity for intimate relationships. *J Pers Soc Psychol* **70**, 1327–1343 (1996).
110. Chapman, A. L., Dixon-Gordon, K. L., Butler, S. M. & Walters, K. N. Emotional reactivity to social rejection versus a frustration induction among persons with borderline personality features. *Personal. Disord. Theory, Res. Treat.* **6**, 88–96 (2015).
111. Guyer, A. E. *et al.* Amygdala and ventrolateral prefrontal cortex function during anticipated peer evaluation in pediatric social anxiety. *Arch Gen Psychiatry* **65**, 1303–1312 (2008).
112. Lenzenweger, M. F., Clarkin, J. F., Fertuck, E. A. & Kernberg, O. F.

Executive neurocognitive functioning and neurobehavioral systems indicators in borderline personality disorder: a preliminary study. *J Pers Disord* **18**, 421–438 (2004).
113. Gardner, D. L., Leibenluft, E., O'Leary, K. M. & Cowdry, R. W. Self-ratings of anger and hostility in borderline personality disorder. *J Nerv Ment Dis* **179**, 157–161 (1991).
114. Rusch, N. *et al.* Shame and implicit self-concept in women with borderline personality disorder. *Am J Psychiatry* **164**, 500–508 (2007).
115. Tomko, R. L., Trull, T. J., Wood, P. K. & Sher, K. J. Characteristics of borderline personality disorder in a community sample: comorbidity, treatment utilization, and general functioning. *J Pers Disord* **28**, 734–750 (2014).
116. First, M. B., Spitzer, R. L., Gibbon, M. & Williams, J. B. W. Structured clinical interview for DSM-IV axis I disorders. *New York New York State Psychiatr. Inst.* (1995).
117. Loranger, A. W., Janca, A. & Sartorius, N. *Assessment and diagnosis of personality disorders: The ICD-10 international personality disorder examination (IPDE)*. (Cambridge University Press, 1997).
118. Jacob, G. A., Ower, N. & Buchholz, A. The role of experiential avoidance, psychopathology, and borderline personality features in experiencing positive emotions: a path analysis. *J Behav Ther Exp Psychiatry* **44**, 61–68 (2013).
119. Taylor, S. E. & Brown, J. D. Illusion and well-being: a social psychological perspective on mental health. *Psychol Bull* **103**, 193 (1988).
120. Korn, C. W., Prehn, K., Park, S. Q., Walter, H. & Heekeren, H. R. Positively biased processing of self-relevant social feedback. *J. Neurosci.* **32**, 16832–16844 (2012).
121. Weiner, B., Kun, A. & Benesh-Weiner, M. The development of mastery, emotions, and morality from an attributional perspective. in *Development of Cognition, Affect and Social Relations. The Minnesota Symposia on Child Psychology* vol. 13 103–130 (Lawrence Erlbaum Hillsdale, NJ, 1980).
122. Tackett, J. L. *et al.* It's Time to Broaden the Replicability Conversation: Thoughts for and From Clinical Psychological Science. *Perspect Psychol Sci* **12**, 742–756 (2017).
123. Ochsner, K. N. & Gross, J. J. The cognitive control of emotion. *Trends Cogn Sci* **9**, 242–249 (2005).
124. Zanarini, M. C., Frankenburg, F. R., Reich, D. B. & Fitzmaurice, G. M. Fluidity of the Subsyndromal Phenomenology of Borderline

Personality Disorder Over 16 Years of Prospective Follow-Up. *Am J Psychiatry* **173**, 688–694 (2016).
125. Stepp, S. D., Pilkonis, P. A., Yaggi, K. E., Morse, J. Q. & Feske, U. Interpersonal and emotional experiences of social interactions in borderline personality disorder. *J Nerv Ment Dis* **197**, 484–491 (2009).
126. Bouchard, S., Sabourin, S., Lussier, Y. & Villeneuve, E. Relationship quality and stability in couples when one partner suffers from borderline personality disorder. *J Marital Fam Ther* **35**, 446–455 (2009).
127. Ross, J. M. & Babcock, J. C. Proactive and reactive violence among intimate partner violent men diagnosed with antisocial and borderline personality disorder. *J. Fam. Violence* **24**, 607–617 (2009).
128. Jeung, H., Schwieren, C. & Herpertz, S. C. Rationality and self-interest as economic-exchange strategy in borderline personality disorder: Game theory, social preferences, and interpersonal behavior. *Neurosci. Biobehav. Rev.* **71**, (2016).
129. Unoka, Z., Seres, I., Aspan, N., Bodi, N. & Keri, S. Trust game reveals restricted interpersonal transactions in patients with borderline personality disorder. *J Pers Disord* **23**, 399–409 (2009).
130. King-Casas, B. *et al.* The rupture and repair of cooperation in borderline personality disorder. *Science (80-.).* **321**, 806–810 (2008).
131. Saunders, K. E., Goodwin, G. M. & Rogers, R. D. Borderline personality disorder, but not euthymic bipolar disorder, is associated with a failure to sustain reciprocal cooperative behaviour: implications for spectrum models of mood disorders. *Psychol Med* **45**, 1591–1600 (2015).
132. Falk, A. & Fischbacher, U. A theory of reciprocity. *Games Econ. Behav.* **54**, 293–315 (2006).
133. Fehr, E. & Schmidt, K. M. A Theory of Fairness, Competition, and Cooperation. *Q. J. Econ.* **114**, 817–868 (1999).
134. Wischniewski, J. & Brune, M. How do people with borderline personality disorder respond to norm violations? Impact of personality factors on economic decision-making. *J Pers Disord* **27**, 531–546 (2013).
135. Polgar, P., Fogd, D., Unoka, Z., Siraly, E. & Csukly, G. Altered social decision making in borderline personality disorder: an Ultimatum Game study. *J Pers Disord* **28**, 841–852 (2014).
136. Cousineau, P. & Young, J. E. [Treatment of borderline personality disorder with the schema-focused approach]. *Sante Ment Que* **22**, 87–105 (1997).

137. De Panfilis, C. Increased punishment behaviors during fair interpersonal exchanges in borderline personality disorder. (2017).
138. Lis, S. Solidarity in the face of chance: Effects of sensitivity to injustice, expectations and anticipated emotions in BPD. *Present. XV. Int. Soc. Study Personal. Disord. Congr. Sept. 26, 2017; Heidelb.* (2017).
139. Unoka, Z. Does the proposer's unfair proposal or fair intention have greater effect on the interaction behavior of patients suffering from borderline personality disorder in the constrained proposer Ultimatum Game? *Present. XV. Int. Soc. Study Personal. Disord. Congr. Sept. 26, 2017; Heidelb.* (2017).
140. Okada, A. & Riedl, A. Inefficiency and social exclusion in a coalition formation game: experimental evidence. *Games Econ. Behav.* **50**, 278–311 (2005).
141. Bock, O., Baetge, I. & Nicklisch, A. hroot: Hamburg registration and organization online tool. *Eur. Econ. Rev.* **71**, 117–120 (2014).
142. Derogatis, L. R. *BSI, Brief Symptom Inventory: administration, scoring & procedures manual*. (National Computer Systems, 1993).
143. Bohus, M. *et al.* The Short Version of the Borderline Symptom List (BSL-23): Development and Initial Data on Psychometric Properties. *Psychopathology* **42**, 32–39 (2009).
144. Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **10**, 171–178 (2007).
145. Blanco, M., Engelmann, D., Koch, A. K. & Normann, H.-T. Belief elicitation in experiments: is there a hedging problem? *Exp. Econ.* **13**, 412–438 (2010).
146. Zanarini, M. C. *et al.* Reported pathological childhood experiences associated with the development of borderline personality disorder. *Am J Psychiatry* **154**, 1101–1106 (1997).
147. Fonagy, P. & Bateman, A. The development of borderline personality disorder--a mentalizing model. *J Pers Disord* **22**, 4–21 (2008).
148. Preissler, S., Dziobek, I., Ritter, K., Heekeren, H. R. & Roepke, S. Social cognition in borderline personality disorder: evidence for disturbed recognition of the emotions, thoughts, and intentions of others. *Front Behav Neurosci* **4**, 182 (2010).
149. Kernberg, O. Borderline conditions and pathological narcissism. *New York Jason Aronson* (1975).
150. Young, Y. E., Klosko, J. S. & Weishaar, M. E. Schema Therapy: A Practitioner's Guide. *New York Guilford Press* (2003).
151. Arntz, A., Klokman, J. & Sieswerda, S. An experimental test of the schema mode model of borderline personality disorder. *J Behav Ther*

Exp Psychiatry **36**, 226–239 (2005).
152. Jeung, H., Walther, S., Korn, C. W., Bertsch, K. & Herpertz, S. C. Emotional responses to receiving peer feedback on opinions in borderline personality disorder. *Personal. Disord. Theory, Res. Treat.* **9**, (2018).
153. Lis, S. *et al.* Borderline Personality Disorder Features and Sensitivity to Injustice. *J Pers Disord* **32**, 192–206 (2018).
154. Preuss, N. *et al.* Inconsistency and social decision making in patients with Borderline Personality Disorder. *Psychiatry Res.* **243**, 115–122 (2016).
155. Andreoni, J. Why free ride?: Strategies and learning in public goods experiments. *J. Public Econ.* **37**, 291–304 (1988).
156. Bateman, A. W. & Krawitz, R. *Borderline personality disorder: an evidence-based guide for generalist mental health professionals*. (Oxford University Press, 2013).
157. Linehan, M. M. *et al.* Dialectical behavior therapy versus comprehensive validation therapy plus 12-step for the treatment of opioid dependent women meeting criteria for borderline personality disorder. *Drug Alcohol Depend* **67**, 13–26 (2002).
158. Rizvi, S. L., Dimeff, L. A., Skutch, J., Carroll, D. & Linehan, M. M. A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. *Behav Ther* **42**, 589–600 (2011).
159. Feliu-Soler, A. *et al.* Fostering Self-Compassion and Loving-Kindness in Patients With Borderline Personality Disorder: A Randomized Pilot Study. *Clin. Psychol. Psychother.* **24**, 278–286 (2017).
160. Bowen, K. L., Morgan, J. E., Moore, S. C. & Van Goozen, S. H. M. Young Offenders' Emotion Recognition Dysfunction Across Emotion Intensities: Explaining Variation Using Psychopathic Traits, Conduct Disorder and Offense Severity. *J. Psychopathol. Behav. Assess.* **36**, 60–73 (2014).
161. Shearer, S. L. Phenomenology of self-injury among inpatient women with borderline personality disorder. *J Nerv Ment Dis* **182**, 524–526 (1994).
162. Fairchild, G., Van Goozen, S. H. M., Calder, A. J., Stollery, S. J. & Goodyer, I. M. Deficits in facial expression recognition in male adolescents with early-onset or adolescence-onset conduct disorder. *J. Child Psychol. Psychiatry.* **50**, 627–636 (2009).
163. Jusyte, A. & Schönenberg, M. Impaired social cognition in violent offenders: perceptual deficit or cognitive bias? *Eur. Arch. Psychiatry*

Clin. Neurosci. **267**, 257–266 (2017).
164. Schönenberg, M., Mayer, S. V., Christian, S., Louis, K. & Jusyte, A. Facial Affect Recognition in Violent and Nonviolent Antisocial Behavior Subtypes. *J. Pers. Disord.* **30**, 708–719 (2016).
165. Corden, B., Critchley, H. D., Skuse, D. & Dolan, R. J. Fear recognition ability predicts differences in social cognitive and neural functioning in men. *J. Cogn. Neurosci.* **18**, 889–897 (2006).
166. Smith, M. L., Cottrell, G. W., Gosselin, F. & Schyns, P. G. Transmitting and decoding facial expressions. *Psychol. Sci.* **16**, 184–189 (2005).
167. Boll, S. & Gamer, M. Psychopathic traits affect the visual exploration of facial expressions. *Biol. Psychol.* **117**, 194–201 (2016).
168. Bartz, J. A., Zaki, J., Bolger, N. & Ochsner, K. N. Social effects of oxytocin in humans: context and person matter. *Trends Cogn. Sci.* **15**, 301–309 (2011).
169. MacDonald, K. & MacDonald, T. M. The peptide that binds: a systematic review of oxytocin and its prosocial effects in humans. *Harv. Rev. Psychiatry* **18**, 1–21 (2010).
170. Shamay-Tsoory, S. G. & Abu-Akel, A. The Social Salience Hypothesis of Oxytocin. *Biological Psychiatry* vol. 79 194–202 (2016).
171. Shahrestani, S., Kemp, A. H. & Guastella, A. J. The impact of a single administration of intranasal oxytocin on the recognition of basic emotions in humans: A meta-analysis. *Neuropsychopharmacology* **38**, 1929–1936 (2013).
172. Di Simplicio, M., Massey-Chase, R., Cowen, P. J. & Harmer, C. J. Oxytocin enhances processing of positive versus negative emotional information in healthy male volunteers. *J. Psychopharmacol.* **23**, 241–248 (2009).
173. Domes, G., Kumbier, E., Heinrichs, M. & Herpertz, S. C. Oxytocin promotes facial emotion recognition and amygdala reactivity in adults with asperger syndrome. *Neuropsychopharmacology* **39**, 698–706 (2014).
174. Guastella, A. J. *et al.* The effects of a course of intranasal oxytocin on social behaviors in youth diagnosed with autism spectrum disorders: a randomized controlled trial. *J. Child Psychol. Psychiatry.* **56**, 444–452 (2015).
175. Lischke, A. *et al.* Intranasal oxytocin enhances emotion recognition from dynamic facial expressions and leaves eye-gaze unaffected. *Psychoneuroendocrinology* **37**, 475–481 (2012).
176. Marsh, A. A., Yu, H. H., Pine, D. S. & Blair, R. J. R. Oxytocin improves specific recognition of positive facial expressions.

Psychopharmacology (Berl). **209**, 225–232 (2010).
177. Schulze, L. *et al.* Oxytocin increases recognition of masked emotional faces. *Psychoneuroendocrinology* **36**, 1378–1382 (2011).
178. Gamer, M., Zurowski, B. & Büchel, C. Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9400–9405 (2010).
179. Domes, G. *et al.* Intranasal oxytocin increases covert attention to positive social cues. *Psychol. Med.* **43**, 1747–1753 (2013).
180. Domes, G., Steiner, A., Porges, S. W. & Heinrichs, M. Oxytocin differentially modulates eye gaze to naturalistic social signals of happiness and anger. *Psychoneuroendocrinology* **38**, 1198–1202 (2013).
181. Hubble, K. *et al.* Oxytocin Reduces Face Processing Time but Leaves Recognition Accuracy and Eye-Gaze Unaffected. *J. Int. Neuropsychol. Soc.* **23**, 23–33 (2017).
182. Leng, G. & Ludwig, M. Intranasal Oxytocin: Myths and Delusions. *Biol. Psychiatry* **79**, 243–250 (2016).
183. Nave, G., Camerer, C. & McCullough, M. Does Oxytocin Increase Trust in Humans? A Critical Review of Research. *Perspect. Psychol. Sci.* **10**, 772–789 (2015).
184. Walum, H., Waldman, I. D. & Young, L. J. Statistical and Methodological Considerations for the Interpretation of Intranasal Oxytocin Studies. *Biol. Psychiatry* **79**, 251–257 (2016).
185. Meyer-Lindenberg, A., Domes, G., Kirsch, P. & Heinrichs, M. Oxytocin and vasopressin in the human brain: Social neuropeptides for translational medicine. *Nature Reviews Neuroscience* vol. 12 524–538 (2011).
186. Blair, R. J. R. Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *J. Neurol. Neurosurg. Psychiatry* **71**, 727–731 (2001).
187. Paulhus, D. L.; Neumann, C.S.; Hare, R. D. . Manual for the Self-Report Psychopathy (SRP) scale: 4th edition. *Toronto Multi-Health Syst.* (2016).
188. Spielberger, C. D. Manual for the state-trait anger expression scale (STAXI). *Psychol. Assess. Resour. Odessa* (1988).
189. Coccaro, E. F., Berman, M. E. & Kavoussi, R. J. Assessment of life history of aggression: development and psychometric characteristics. *Psychiatry Res.* **73**, 147–157 (1997).
190. Raine, A. *et al.* The Reactive-Proactive Aggression Questionnaire: Differential Correlates of Reactive and Proactive Aggression in

Adolescent Boys. *Aggress. Behav.* **32**, 159–171 (2006).
191. Born, J. *et al.* Sniffing neuropeptides: a transnasal approach to the human brain. *Nat. Neurosci.* **5**, 514–516 (2002).
192. Spengler, F. B. *et al.* Kinetics and Dose Dependency of Intranasal Oxytocin Effects on Amygdala Reactivity. *Biol. Psychiatry* **82**, 885–894 (2017).
193. Steyer, R., Schwenkmezger, P., Notz, P. & Eid, M. MDBF–Mehrdimensionaler Befindlichkeitsfragebogen. *Hogrefe, Göttingen, Deutschl.* (1997).
194. Tottenham, N. NimStim Face Stimulus Set. Research Network on Early Experience and Brain Development. *https://www.macbrain.org* (2017).
195. Ekman, P. & Friesen, W. V. Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* **17**, 124–129 (1971).
196. Lundqvist, Daniel, Anders Flykt, and A. Ö. Karolinska directed emotional faces. *Cogn. Emot.* (1998).
197. Ebner, N. C., Riediger, M. & Lindenberger, U. FACES--a database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav. Res. Methods* **42**, 351–362 (2010).
198. Striepens, N. *et al.* Elevated cerebrospinal fluid and blood concentrations of oxytocin following its intranasal administration in humans. *Sci. Rep.* **3**, (2013).
199. Blair, R. J. R. A cognitive developmental approach to mortality: investigating the psychopath. *Cognition* **57**, 1–29 (1995).
200. Blair, R. J. R. Neuroimaging of psychopathy and antisocial behavior: a targeted review. *Curr. Psychiatry Rep.* **12**, 76–82 (2010).
201. Woodbury-Smith, M. R. *et al.* A case-control study of offenders with high functioning autistic spectrum disorders. *http://dx.doi.org/10.1080/14789940500302554* **16**, 747–763 (2007).
202. Marissen, M. A. E., Deen, M. L. & Franken, I. H. A. Disturbed emotion recognition in patients with narcissistic personality disorder. *Psychiatry Res.* **198**, 269–273 (2012).
203. Blair, R. J. R. The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience* vol. 14 786–799 (2013).
204. Gamer, M. & Büchel, C. Amygdala activation predicts gaze toward fearful eyes. *J. Neurosci.* **29**, 9123–9126 (2009).
205. Marshall, A. D. & Holtzworth-Munroe, A. Varying forms of husband sexual aggression: predictors and subgroup differences. *J. Fam. Psychol.* **16**, 286–296 (2002).
206. Morgan, J. E., Bowen, K. L., Moore, S. C. & van Goozen, S. H. M. The relationship between reward and punishment sensitivity and

antisocial behavior in male adolescents. *Pers. Individ. Dif.* **63**, 122–127 (2014).
207. Roose, A., Bijttebier, P., Claes, L. & Lilienfeld, S. O. Psychopathic traits in adolescence: Associations with the revised Reinforcement Sensitivity Theory systems. *Pers. Individ. Dif.* **50**, 201–205 (2011).
208. Young, L. J. & Barrett, C. E. Neuroscience. Can oxytocin treat autism? *Science* **347**, 825–856 (2015).
209. Gao, S. *et al.* Oxytocin, the peptide that bonds the sexes also divides them. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7650–7654 (2016).
210. Bartz, J. A. *et al.* Oxytocin selectively improves empathic accuracy. *Psychol. Sci.* **21**, 1426–1428 (2010).
211. Ne'eman, R., Perach-Barzilay, N., Fischer-Shofty, M., Atias, A. & Shamay-Tsoory, S. G. Intranasal administration of oxytocin increases human aggressive behavior. *Horm. Behav.* **80**, 125–131 (2016).
212. Turner, D., Sebastian, A. & Tüscher, O. Impulsivity and Cluster B Personality Disorders. *Curr. Psychiatry Rep.* **19**, (2017).
213. Herpertz, S. C. & Bertsch, K. A New Perspective on the Pathophysiology of Borderline Personality Disorder: A Model of the Role of Oxytocin. *Am. J. Psychiatry* **172**, 840–851 (2015).
214. Lischke, A. *et al.* Oxytocin increases amygdala reactivity to threatening scenes in females. *Psychoneuroendocrinology* **37**, 1431–1438 (2012).
215. Wei, D. *et al.* Endocannabinoid signaling mediates oxytocin-driven social reward. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14084–14089 (2015).
216. Blair, R. J. R. Facial expressions, their communicatory functions and neuro-cognitive substrates. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **358**, 561–572 (2003).
217. De Dreu, C. K. & Kret, M. E. Oxytocin Conditions Intergroup Relations Through Upregulated In-Group Empathy, Cooperation, Conformity, and Defense. *Biol Psychiatry* **79**, 165–173 (2016).
218. Dadds, M. R. *et al.* Nasal oxytocin for social deficits in childhood autism: a randomized controlled trial. *J. Autism Dev. Disord.* **44**, 521–531 (2014).
219. Huang, H. *et al.* Chronic and acute intranasal oxytocin produce divergent social effects in mice. *Neuropsychopharmacology* **39**, 1102–1114 (2014).
220. Oya, K., Matsuda, Y., Matsunaga, S., Kishi, T. & Iwata, N. Efficacy and safety of oxytocin augmentation therapy for schizophrenia: an updated systematic review and meta-analysis of randomized, placebo-controlled trials. *Eur. Arch. Psychiatry Clin. Neurosci.* **266**, 439–450 (2016).

221. Yatawara, C. J., Einfeld, S. L., Hickie, I. B., Davenport, T. A. & Guastella, A. J. The effect of oxytocin nasal spray on social interaction deficits observed in young children with autism: a randomized clinical crossover trial. *Mol. Psychiatry* **21**, 1225–1231 (2016).
222. Hunt, V., Layton, D. & Prince J A N U A, S. Why diversity matters. (2015).
223. Badura, K. L., Grijalva, E., Newman, D. A., Yan, T. T. & Jeon, G. Gender and leadership emergence: A meta-analysis and explanatory model. *Wiley Online Libr.* **71**, 335–367 (2018).
224. The female CEOs on the Fortune 500 just broke three all-time records | Fortune. https://fortune.com/2021/06/02/female-ceos-fortune-500-2021-women-ceo-list-roz-brewer-walgreens-karen-lynch-cvs-thasunda-brown-duckett-tiaa/.
225. How Many Fortune 500 CEOs Are Women? And Why So Few? - The Quantic Blog. https://quantic.edu/blog/2021/12/06/how-many-fortune-500-ceos-are-women/.
226. Global report: Women in Business and Management: Gaining momentum. https://www.ilo.org/global/publications/ilo-bookstore/order-online/books/WCMS_316450/lang--en/index.htm.
227. Lord, R., Vader, C. De, psychology, G. A.-J. of applied & 1986, undefined. A meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *psycnet.apa.org*.
228. Ensari, N., Riggio, R. E., Christian, J. & Carslaw, G. Who emerges as a leader? Meta-analyses of individual differences as predictors of leadership emergence. *Pers. Individ. Dif.* **51**, 532–536 (2011).
229. Judge, T. A., Piccolo, R. F. & Kosalka, T. The bright and dark sides of leader traits: A review and theoretical extension of the leader trait paradigm. *Leadersh. Q.* **20**, 855–875 (2009).
230. O'Reilly, C., Review, J. C.-C. M. & 2020, undefined. Transformational leader or narcissist? How grandiose narcissists can create and destroy organizations and institutions. *journals.sagepub.com* **62**, 5–27 (2020).
231. Chiorri, C., Garofalo, C., psychology, P. V. & 2019, undefined. Does the dark triad manifest similarly in men and women? Measurement invariance of the dirty dozen across sex. *Springer*.
232. Muris, P., Merckelbach, H., … H. O.-P. on & 2017, undefined. The malevolent side of human nature: A meta-analysis and critical review of the literature on the dark triad (narcissism, Machiavellianism, and psychopathy). *journals.sagepub.com* **12**, 183–204 (2017).
233. Crysel, L. C., Crosier, B. S. & Webster, G. D. The Dark Triad and risk

behavior q. *Elsevier* (2012) doi:10.1016/j.paid.2012.07.029.
234. Babcock, L., Recalde, M. P., Vesterlund, L. & Weingart, L. Gender differences in accepting and receiving requests for tasks with low promotability. *Am. Econ. Rev.* **107**, 714–747 (2017).
235. Murphy, R. O., Ackermann, K. A. & Handgraaf, M. J. J. Measuring Social Value Orientation. *Judgm. Decis. Mak.* **6**, 771–781 (2011).
236. Armin, F., Becker, A., Dohmen, T. J., Huffman, D. & Sunde, U. The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. *SSRN Electron. J.* (2016) doi:10.2139/ssrn.2725874.
237. Duch, M. L., Grossmann, M. R. P. & Lauer, T. z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the Internet. *J. Behav. Exp. Financ.* **28**, 100400 (2020).
238. Jones, D. N. & Paulhus, D. L. Introducing the Short Dark Triad (SD3). *Assessment* **21**, 28–41 (2014).
239. Davis, M. H. A multidimensional approach to individual differences in empathy. *JSAS Cat. Sel. Doc. Psychol.* **10**, 85 (1980).
240. Balliet, D., Li, N. P., Macfarlan, S. J. & Van Vugt, M. Sex Differences in Cooperation: A Meta-Analytic Review of Social Dilemmas. *Psychol. Bull.* **137**, 881–909 (2011).
241. Capraro, V. Women are slightly more cooperative than men (in one-shot Prisoner's dilemma games played online). (2018).
242. Wai, M. & Tiliopoulos, N. The affective and cognitive empathic nature of the dark triad of personality. *Elsevier* (2012) doi:10.1016/j.paid.2012.01.008.
243. Hosseini-Kamkar, N. & Bruce Morton, J. Sex differences in self-regulation: an evolutionary perspective. *Front. Neurosci.* **8**, (2014).
244. Silverman, I. W. Gender Differences in Delay of Gratification: A Meta-Analysis. *Sex Roles* **49**, 451–463 (2003).
245. Jonason, P. K., Koenig, B. L. & Tost, J. Living a Fast Life The Dark Triad and Life History Theory. *Springer* doi:10.1007/s12110-010-9102-4.
246. Heilman, M. E. & Chen, J. J. Same behavior, different consequences: Reactions to men's and women's altruistic citizenship behavior. *J. Appl. Psychol.* **90**, 431–441 (2005).
247. Jonason, P. K. & Webster, G. D. The dirty dozen: A concise measure of the dark triad. *Psychol. Assess.* **22**, 420–432 (2010).
248. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science (80-.).* **211**, 1390–1396 (1981).
249. Flood, M. M. Some experimental games. *Manage. Sci.* **5**, 5–26 (1958).
250. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image

scoring. *Nature* **393**, 573–577 (1998).
251. Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. The evolution of altruistic punishment. *Proc. Natl. Acad. Sci.* **100**, 3531–3535 (2003).
252. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
253. Domes, G. *et al.* Effects of intranasal oxytocin on emotional face processing in women. *Psychoneuroendocrinology* **35**, 83–93 (2010).
254. Ma, Y. *et al.* Intranasal oxytocin attenuates insula activity in response to dynamic angry faces. *Biol. Psychol.* **157**, (2020).
255. Domes, G. *et al.* Oxytocin Attenuates Amygdala Responses to Emotional Faces Regardless of Valence. *Biol. Psychiatry* **62**, 1187–1190 (2007).
256. Lobbestael, J., Cima, M. & Arntz, A. The relationship between adult reactive and proactive aggression, hostile interpretation bias, and antisocial personality disorder. *J. Pers. Disord.* **27**, 53–66 (2013).
257. Modestin, J. Quality of interpersonal relationships: the most characteristic DSM-III BPD criterion. *Compr Psychiatry* **28**, 397–402 (1987).
258. Zanarini, M. C., Gunderson, J. G., Frankenburg, F. R. & Chauncey, D. L. Discriminating borderline personality disorder from other axis II disorders. *Am J Psychiatry* **147**, 161–167 (1990).
259. Gunderson, J. G. Disturbed relationships as a phenotype for borderline personality disorder. *Am J Psychiatry* **164**, 1637–1640 (2007).
260. de Groot, E. R., Verheul, R. & Trijsburg, R. W. An integrative perspective on psychotherapeutic treatments for borderline personality disorder. *J Pers Disord* **22**, 332–352 (2008).
261. Stoffers, J. M. *et al.* Psychological therapies for people with borderline personality disorder. *Cochrane Database Syst Rev* **8**, doi: 10.1002/14651858 (2012).
262. Benjamin, L. S. & Wonderlich, S. A. Social perceptions and borderline personality disorder: the relation to mood disorders. *J Abnorm Psychol* **103**, 610–624 (1994).
263. Black, D. W. *et al.* Attitudes toward borderline personality disorder: a survey of 706 mental health clinicians. *CNS Spectr* (2011).
264. Vaillant, G. E. The beginning of wisdom is never calling a patient a borderline; or, the clinical management of immature defenses in the treatment of individuals with personality disorders. *J Psychother Pr. Res* **1**, 117–134 (1992).
265. Lieb, K., Vollm, B., Rucker, G., Timmer, A. & Stoffers, J. M.

Pharmacotherapy for borderline personality disorder: Cochrane systematic review of randomised trials. *Br J Psychiatry* **196**, 4–12 (2010).
266. Zanarini, M. C. Update on pharmacotherapy of borderline personality disorder. *Curr Psychiatry Rep* **6**, 66–70 (2004).
267. NICE Clinical Guidelines. Antisocial behaviour and conduct disorders in children and young people: recognition and management. *Antisocial Behav. Conduct Disord. Child. young people Recognit. Manag.* (2017).
268. Brazil, I. A., van Dongen, J. D. M., Maes, J. H. R., Mars, R. B. & Baskin-Sommers, A. R. Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience and Biobehavioral Reviews* vol. 91 259–277 (2018).
269. Soeteman, D. I., Roijen, L. H. Van, Verheul, R. & Busschbach, J. J. V. The economic burden of personality disorders in mental health care. *J. Clin. Psychiatry* **69**, 259–265 (2008).
270. Wunsch, E. M., Kliem, S. & Kröger, C. Population-based cost-offset estimation for the treatment of borderline personality disorder: projected costs in a currently running, ideal health system. *Behav. Res. Ther.* **60**, 1–7 (2014).
271. Gatner, D. T., Douglas, K. S., Almond, M. F. E., Hart, S. D. & Kropp, P. R. How much does that cost? Examining the economic costs of crime in North America attributable to people with psychopathic personality disorder. *Personal. Disord.* (2022) doi:10.1037/PER0000575.
272. Judge, T. A., LePine, J. A. & Rich, B. L. Loving yourself abundantly: Relationship of the narcissistic personality to self- and other perceptions of workplace deviance, leadership, and task and contextual performance. *J. Appl. Psychol.* **91**, 762–776 (2006).
273. Baughman, H. M., Dearing, S., Giammarco, E. & Vernon, P. A. Relationships between bullying behaviours and the Dark Triad: A study with adults. *Pers. Individ. Dif.* **52**, 571–575 (2012).
274. Makowsky, M. D. & Smaldino, P. E. The evolution of power and the divergence of cooperative norms. *J. Econ. Behav. Organ.* **126**, 75–88 (2016).
275. Simonsen, S. *et al.* European guidelines for personality disorders: past, present and future. *Borderline Personal. Disord. Emot. Dysregulation* **6**, 9 (2019).
276. No Title. in *Borderline Personality Disorder: Treatment and Management* (2009).
277. Clinical practice guideline for the management of borderline

personality disorder.
http://www.nhmrc.gov.au/guidelines/publications/mh25 (2013).
278. Zanarini, M. C., Frankenburg, F. R., Hennen, J. & Silk, K. R. Mental health service utilization by borderline personality disorder patients and Axis II comparison subjects followed prospectively for 6 years. *J Clin Psychiatry* **65**, 28–36 (2004).
279. Cristea, I. A. *et al.* Efficacy of psychotherapies for borderline personality disorder: A systematic review and meta-analysis. *JAMA Psychiatry* vol. 74 319–328 (2017).
280. Linehan, M. M. Dialectical behavior therapy for borderline personality disorder. Theory and method. *Bull Menn. Clin* **51**, 261–276 (1987).
281. Bateman, A. W. & Fonagy, P. Mentalization-based treatment of BPD. *Journal of Personality Disorders* vol. 18 36–51 (2004).
282. Yeomans, F., Clarkin, J. & Kernberg, O. A primer of transference-focused psychotherapy for the borderline patient. (2002).
283. Cristea, I. A., Barbui, C. & Cuijpers, P. Reviews and meta-analyses of psychotherapy efficacy for borderline personality disorder - Reply. *JAMA Psychiatry* vol. 74 854–855 (2017).
284. Brakemeier, E. L. & Herpertz, S. C. Innovative psychotherapy research: towards an evidence-based and process-based individualized and modular psychotherapy. *Nervenarzt* vol. 90 1125–1134 (2019).
285. Schnell, K. & Herpertz, S. C. Psychotherapy in psychiatry: The current situation and future directions in Germany. *European Archives of Psychiatry and Clinical Neuroscience* vol. 261 (2011).
286. Schwarzinger, D. & Schuler, H. *TOP. Dark Triad of Personality at Work: Manual*. (Hogrefe Verlag, 2016).
287. Grijalva, E., Harms, P. D., Newman, D. A., Gaddis, B. H. & Fraley, R. C. Narcissism and Leadership: A Meta-Analytic Review of Linear and Nonlinear Relationships. *Pers. Psychol.* **68**, 1–47 (2015).
288. Bertsch, K., Florange, J. & Herpertz, S. C. Understanding Brain Mechanisms of Reactive Aggression. *Curr. Psychiatry Rep.* **22**, 1–16 (2020).
289. Stanley, B. & Siever, L. J. The interpersonal dimension of borderline personality disorder: toward a neuropeptide model. *Am J Psychiatry* **167**, 24–39 (2010).
290. Jack, A., Connelly, J. J. & Morris, J. P. DNA methylation of the oxytocin receptor gene predicts neural response to ambiguous social stimuli. *Front Hum Neurosci* **6**, 280 (2012).
291. Axelrod, R. The Evolution of Cooperation. *Basic Books, New York* (1984).

292. Trivers, R. L. The Evolution of Reciprocal Altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).