Superior 'theory of mind' in borderline personality disorder: An analysis of interaction behavior in a virtual trust game

Nele Franzen [a], Meike Hagenhoff [a], Nina Baer [a], Ariane Schmidt [a], Daniela Mier [a,b], Gebhard Sammer [a], Bernd Gallhofer [a], Peter Kirsch [a,b], Stefanie Lis [a,*]

[a] Centre for Psychiatry, Justus-Liebig University Giessen, Giessen, Germany
[b] Division of Clinical Psychology, Central Institute of Mental Health, Mannheim, and Mannheim School of Medicine, University of Heidelberg, Heidelberg, Germany

ARTICLE INFO

ABSTRACT

To gain further insight into interpersonal dysfunction in Borderline Personality Disorder (BPD) we investigated the effects of emotional cues and the fairness of a social partner on the ability to infer other peoples' intentions in a virtual social exchange. 30 BPD patients and 30 nonpatients were asked to play a multiround trust game with four virtual trustees. The trustees varied in regard to fairness and presence of emotional facial cues which were both linked to repayment ratio. BPD patients adjusted their investment to the fairness of their partner. In contrast, nonpatients disregarded the trustees' fairness in the presence of emotional facial expressions. Both groups performed equally in an emotion recognition task and assessed the trustees' fairness comparably. When the unfair trustee provided emotional cues, BPD patients assessed their own behavior as more fair, while the lack of cues led patients to assess their own behavior as unfair. BPD patients are superior in the attribution of mental states to interaction partners when emotional cues are present. While the emotional expressions of a partner dominated the exchange behavior in nonpatients, BPD patients used the objective fairness of their social counterparts to guide their own behavior despite the existence of emotional cues.

© 2010 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Interpersonal dysfunction constitutes a core symptom domain in Borderline personality disorder (BPD), with seven of nine diagnostic criteria in the DSM-IV-TR referring to this area of functioning (Gunderson, 2007; Bouchard and Sabourin, 2009). In general, prerequisite for successful interactions in social settings is the ability to take intentions, emotions, and beliefs of the social interaction partner into account when making decisions (Frith and Singer, 2008). This ability is referred to as theory of mind (ToM, Premack and Woodruff, 1978) or mentalizing (Frith, 1989). It enables the subjects to predict the next move of a social interaction partner and thus adapt the own behavior to achieve a goal. While general beliefs about unknown interaction partners as for example beliefs on trustworthiness or competence might be determined in advance by stable physical features (Willis and Todorov, 2006; Wilson and Eckel, 2006; Santos and Young, 2008), and information given about social partners (Singer et al., 2004), subjects learn about people by interacting with them. Based on past experiences, subjects do not only adjust their behavior e.g. in relation to the fairness of a partner (Singer et al., 2006), but change their attitude towards their counterparts regarding e.g. their trustworthiness (Bayliss and Tipper, 2006) and attractiveness (Singer et al., 2004), as well as the extent of empathy experienced with another subject (Singer et al., 2006). Thereby, emotional facial expressions which represent a form of social cue have a modulating effect and fulfill a crucial communicatory function (Blair, 2003). Smiling faces for instance, could be shown to increase judgments of generosity and sociability on faces of strangers (Scharlemann et al., 2001; Mehu et al., 2007a,b). Thus, the building and maintenance of social relationships require a complex interplay between various processes.

In search of the causes underlying interpersonal dysfunction in BPD various studies have investigated the contribution of single processes assumed to be involved in this impairment. Most of these studies have focussed upon alterations in emotion recognition (review see Domes et al., 2009). However, results across studies are inconsistent. They point to subtle impairments in basic emotion recognition (Levine et al., 1997; Bland et al., 2004), a negativity bias (Wagner and Linehan, 1999; Donegan et al., 2003; Dyck et al., 2009; Guitart-Masip et al., 2009), but also to a heightened sensitivity for the detection of negative emotions (Lynch et al., 2006), or no alterations in emotion recognition at all (Wagner and Linehan, 1999; Minzenberg et al., 2006; Domes et al.,

2008). A study by Minzenberg et al. (2006) suggests that impairments in BPD might become apparent if an integration of information is required as e.g. in the combination of facial and prosodic information to recognize an emotional state. Several studies point to alterations of cerebral correlates of emotion recognition abilities in BPD (Donegan et al., 2003; Schmahl and Bremner, 2006; Minzenberg et al., 2007; Koenigsberg et al., 2009; Marissen et al., 2010).

In contrast to other psychiatric disorders involving impairments in social functioning as schizophrenia or autism, less attention has been directed to deficits in mentalizing. This seems surprising because deficits in mentalizing play an essential role in therapeutical approaches to BPD. In the mentalization-based therapy for instance, the ability to mentalize is considered unstable during emotional arousal and the therapy aims towards an improvement of the patients' abilities to understand their own and other mental states (Fonagy and Bateman, 2008). Only a few recent studies addressed the ability of BPD patients to attribute mental states to other people (Arntz et al., 2009; Fertuck et al., 2009; Harari et al., 2010; Ghiassi et al., 2010). The majority of these indicate an equal or superior mentalization ability in BPD. Two studies homogenously point towards a superior ability of BPD patients to identify mental states. With help of the 'Reading the mind in the eyes test' Fertuck et al. (2009) showed an enhanced sensitivity in BPD when attributing a mental state based on information derived from pictures portraying the eye region of the face. Arntz et al. (2009) applied the Happe Test which requires subjects to infer other subjects' thoughts, feelings, and intentions in complex social situations described in several stories that involve double bluff, mistakes, persuasion, and white lies. BPD patients performed superior to nonpatients when performance was controlled for IQ. In contrast, deficits in ToM could be observed in both of these tests in patients with anorexia nervosa as could be shown in a recent study by Russell et al. (2009). However, Harari et al. (2010) suggest that ToM abilities are impaired in BPD patients at least in a subdomain of ToM abilities. Reduced performance became obvious in cognitive ToM, i.e. the ability to make inferences regarding other people's beliefs, but not in affective ToM, i.e. the ability to make inferences regarding other peoples' emotions. These results contradict the results of Arntz et al. (2009) in that the latter found superior performance for BPD patients in the Happe Test that can be regarded as a cognitive ToM task which requires "thinking about thinking" and deductive reasoning skills (Russell et al., 2009). Beyond that, these deficits could not be found in a study by Ghiassi et al. (2010) who applied the MSAT, a test of cognitive mentalizing skills. They found a comparable performance in BPD and nonpatients during the task in which scenes of cartoon picture stories about social interactions had to be sorted. Additionally, BPD patients and nonpatients assessed the beliefs, intentions, false beliefs, as well as deception and reciprocity of the characters involved in these stories in a comparable manner.

To gain further insight into the factors contributing to alterations in interpersonal functioning in BPD, we chose the alternative approach to analyze processes of mentalizing in a simulated social interaction situation. It was noticed only recently that exchange games as the trust-game that are well established in behavioral economy might constitute a powerful tool to study impairments in social interactions in psychiatric disorders (Meyer-Lindenberg, 2008; Seres et al., 2009). Although these paradigms are mostly embedded theoretically in the context of social decision making, it can be assumed that the assessment of the partner's intentions and a prediction of his future behavior, i.e. the building of a theory of mind, contributes to the decision. Thus, Frith and Singer (2008) state 'mentalizing is important for these economic games" (p 3878) and 'the ability to mentalize may help to determine what the next moves or intentions of the other players might be" (p 3881). Studies which show that the ability to mentalize, as measured by traditonal ToM tasks, is related to cooperation and fair behavior in exchange games, support this view (Sally and Hill, 2006).

In contrast to traditional ToM tasks, exchange games allow to analyze the behavior of subjects directly engaged in a social interaction. Thereby

they might have a higher ecological validity compared to the majority of traditional ToM tasks which require the passive observation and interpretation of interactions of others e.g. in cartoons (Brunet et al., 2003), simple animations (Castelli et al., 2000), or jokes (Langdon and Ward, 2009).

Additionally, exchange games allow not only the measurement of the influences of explicit controlled processes available to verbalisation, but also the measurement of the influences of automatic implicit processes that affect action unconsciously (Frith and Frith, 2008). Beyond that, exchange games offer the opportunity to experimentally manipulate single factors in order to analyze their contribution to alterations in social interaction.

So far, two studies applied the trust game in BPD and both point to alterations in the behavior of BPD patients (King-Casas et al., 2008; Unoka et al., 2009). If an interaction partner transfered only a small investment during the game, BPD patients reacted with an equally low repayment (King-Casas et al., 2008) while nonpatients showed a more generous behavior, i.e. they often returned a high sum regardlessly. In the course of a 10-round game these differences in behavior lead to a break down in cooperation in those exchanges in which BPD patients were involved. King-Casas et al. hypothesized that the alterations in the behavior of BPD patients is a consequence of a lack of perception of violation of social norms that might be linked to negative expectations and negative evaluative biases of social partners. Unoka et al. (2009) could show that BPD patients fail to develop trust in the course of a 5-round trust game during which they did not receive feedback about their interaction partner's behavior: While nonpatients increased their investments over the course of the game, BPD patients did not change their behavior. These data suggest that exchange games might indeed be useful to depict those alterations in interaction behavior that might underly deficits in social functioning.

The aim of the present study was to analyze which factors might contribute to alterations in mental state attribution underlying decision making during social interactions. We were especially interested in whether the fairness of a social partner and the existence of emotional cues lead to deviations in the investment behavior of BPD patients compared to nonpatients. To control the behavior of an interaction partner we chose the approach of a simulated interaction that allows to experimentally manipulate the characteristics of the exchange. In a multiround trust game played with several virtual partners, we manipulated the interaction partner's fairness as well as the emotional facial expression that allowed to infer upon the partner's intentions within an individual exchange round. There are several variables that can be assumed to be impaired in BPD and that might contribute to alterations in the social interaction behavior of BPD patients in the exchange game of the present study. As described above, several studies indicate impairments in the recognition of emotions in faces. We hypothesized that these impairments might result in a reduced efficiency to use emotional expressions as cues in order to predict the interaction partner's intentions and thus lead to an inability in BPD patients to adjust their investment behavior accordingly. To identify deficits in emotion recognition and control their possible contribution to alterations in social interaction behavior in BPD, we measured emotion recognition capabilities in an additionally applied emotion recognition task. Another variable that might affect interaction behavior in a trust game is the manner in which subjects experience the fairness of others, i.e. the applied social norms. King-Casas et al. (2008) assumed that BPD patients differ from nonpatients in that they apply different social norms to judge their social interaction partner's behavior. We hypothesized that BPD patients differ from nonpatients in the assessment of their partner's fairness and consequently differ in the adjustment of their interaction behavior. To identify alterations in social norms and control their possible contribution to alterations in social inter-action behavior in BPD, we measured the assessment of the fairness of the social partners behavior as an indirect measure of social norm application.