"Feeling invisible": Individuals with borderline personality disorder underestimate the transparency of their emotions

Celine De Meulemeester [1], Benedicte Lowyck [2], Bart Boets[3], Stephanie Van der Donck[3], Yannic Verhaest[2], and Patrick Luyten [1,4]

[1] Faculty of Psychology and Educational Sciences, KU Leuven, Belgium

[2] University Psychiatric Hospital UPC KU Leuven, Campus Kortenberg, and Department of Neurosciences, Faculty of Medicine, KU Leuven, Belgium

[3] Center for Developmental Psychiatry, Department of Neurosciences, KU Leuven, Belgium

[4] Research Department of Clinical, Educational and Health Psychology, University College London, UK

Author Note

Conflict of interest: P.L. has been involved in the development, training and dissemination of mentalization-based treatments.

Funding: C.D.M. is supported by a PhD fellowship from the Research Foundation – Flanders (FWO) [11B9718N]; B.B has received the Research Foundation – Flanders (FWO) Excellence of Science grant [G0E8718N; HUMVISCAT] and an Internal KU Leuven grant (C14/17/102); S.V.D.'s work is funded by an Internal KU Leuven grant (PDM/20/170).

Correspondence concerning this article should be addressed to Celine De Meulemeester, Faculty of Psychology and Educational Sciences, Tiensestraat 102 pobox 3722, 3000 Leuven, Belgium. E-mail: celine.demeulemeester@kuleuven.be. ORCID: 0000-0003-1608-5304

Abstract

The present study investigated transparency estimation, that is, the ability to estimate how observable one's emotions are, in patients diagnosed with borderline personality disorder (BPD) (n = 35) and healthy controls (HCs; n = 35). Participants watched emotionally evocative video clips and estimated the transparency of their own emotional experience while watching the clip. Facial expression coding software (FaceReader) quantified their objective transparency. BPD patients felt significantly less transparent than HCs, but there were no differences in objective transparency. BPD patients tended to underestimate the transparency of their emotions compared HCs, who in turn overestimated their transparency. This suggests that BPD patients expect that others will not know how they feel, irrespective of how observable their emotions actually are. We link these findings to low emotional awareness and a history of emotional invalidation in BPD, and discuss their impact on BPD patients' social functioning.

Keywords: borderline personality disorder, transparency estimation, facial expressions, mentalizing, emotional arousal

Borderline personality disorder (BPD) is a severe psychiatric condition characterized by intense fluctuations in mood, identity disturbance, interpersonal problems, impulsivity, and self-harming behaviors (Gunderson et al., 2018). The lifetime prevalence of BPD is approximately 6% in the general population (Grant et al., 2008), and about 22%–26% of outpatients in psychiatric settings have a diagnosis of BPD (Korzekwa et al., 2008). Several symptoms of BPD have been related to impairments in mentalizing, that is, the capacity to reflect on one's own and others' mental states (Fonagy & Luyten, 2009). With regard to mentalizing about their own mental states, BPD patients have been found to have low emotional awareness, also called alexithymia (Derks et al., 2017), and a lack of self-insight (Morey, 2014).

A recent study of adolescent BPD patients found that physiological and behavioral measures of their stress were better predicted by an independent observer's rating of stress than by their own self-estimation, with self-perceived stress being lower than other-perceived stress, whereas adolescents without BPD showed the opposite pattern (Bourvis et al., 2020). Adolescent BPD patients' perceptions of their own stress thus seemed to be less aligned with physiological measures of stress compared to others' perceptions, which may indicate a problem with self-reflection. This observation is in line with previous studies finding a dissociation between blunted subjectively experienced emotion and heightened objectively measured emotional responses in BPD patients (for a review, see Perez-Rodriguez et al. (2018)). This reduced awareness of their own emotions may extend to a reduced awareness of how their emotional experience is perceived by others.

BPD patients may have problems with transparency estimation, that is, the ability to estimate the extent to which one's own emotions are observable to others. Previous studies in the general population have typically found a tendency to overestimate the transparency of one's emotional states (Savitsky & Gilovich, 2003; Van Boven et al., 2003; Vorauer et al.,

1999). This is hypothesized to result from a tendency for people to "adjust insufficiently from the anchor of their own phenomenological experience" when taking another person's perspective on the self (T. Gilovich et al., 1998) and this can lead to interpersonal misunderstandings (Cameron & Vorauer, 2008). Considering that individuals with BPD seem less aware of their own emotional states (Bourvis et al., 2020; Derks et al., 2017; New et al., 2012; Perez-Rodriguez et al., 2018) and have more difficulty reflecting on what others are thinking (Fonagy & Luyten, 2009; Németh et al., 2018), transparency estimation errors may be even larger in this population.

Indeed, a larger discrepancy was found between transparency estimations and the objective transparency of those emotions in nonclinical female students with elevated BPD features compared to participants without BPD features (De Meulemeester et al., 2021). This suggests both larger over- and underestimation errors associated with BPD, and these errors were positively related to attachment anxiety and mentalizing impairments. An impaired capacity for transparency estimation may render individuals with BPD less able to gauge how they are perceived by others, which may negatively impact their social interactions. As social dysfunction is a highly debilitating and slow-to-remit symptom of BPD (Lis & Bohus, 2013), it is crucial to further understand the process of transparency estimation in BPD, and how it may relate to these patients' interpersonal problems and other symptoms.

The De Meulemeester et al. (2021) study included a sample of university students that was homogenous in age, gender and education level, which does not represent the total population of individuals displaying BPD features. Therefore, the present study aimed to replicate these findings in a more diverse sample of clinically diagnosed BPD patients enrolled in specialized treatment, as compared to matched healthy controls (HCs). We expect to find transparency estimation errors in the BPD group that are larger in magnitude compared to the De Meulemeester et al. (2021) study due to the larger severity of BPD symptoms in the

present sample. Furthermore, the present study aims to clarify how clinical features such as mentalizing impairments, attachment insecurity, childhood trauma history, interpersonal problems and identity diffusion relate to transparency underestimation, overestimation, or both types of error.

Several mechanisms may determine the type of transparency estimation errors, that is, either over- or underestimation, observed in BPD patients. First, the degree of subjective emotional arousal reported by participants while estimating the transparency of their emotions may be a crucial determinant of transparency overestimation. Indeed, previous studies have found that heightened self-salience of one's own mental state related to increased transparency overestimation (Vorauer et al., 1999), and increased subjective arousal (SAR) during a transparency task related to increased transparency overestimation errors in participants with and without BPD features (De Meulemeester et al., 2021). It may be that individuals reporting higher emotional arousal are highly aware of their own emotions and overestimate the extent to which they are observable to others. This may be even more the case for individuals with BPD, who have been found to have difficulty with third-person perspective taking (Colle et al., 2018; New et al., 2012; Petersen et al., 2016). Reporting low emotional arousal, however, may relate to BPD patients' difficulty in identifying and describing their own emotional states (Derks et al., 2017), as well as their reduced awareness of the bodily signals that may communicate these emotions to others (i.e., impaired interoception) (Back & Bertsch, 2020). As a result, BPD patients may be unaware of their own emotions, erroneously assuming that they are not transparent to others, while in reality their emotions may be observable through various bodily signals (e.g., facial expressions, skin temperature, muscle tension).

Second, the particular type of mentalizing impairment exhibited by a BPD patient may in part determine whether transparency errors are expressed as over- or underestimation. On

the one hand, BPD patients have been found to be characterized by hypermentalizing, that is, a tendency to be extremely certain about mental states and to create excessive and overly complex models of the mind (Sharp & Vanwoerden, 2015). This certainty in knowing one's own mental states and those of others may lead to a failure to recognize that it is impossible to actually read another person's mind and that mental states are in fact opaque unless they are communicated, and may thus result in transparency overestimation. On the other hand, hypomentalizing, or extreme uncertainty about the mental states of self and others (Luyten et al., 2020), may reflect a general lack of awareness of mental states, and may lead to both types of transparency estimation error. Indeed, this pattern of associations was found in nonclinical participants high and low in BPD features (De Meulemeester et al., 2021), but this finding needs to be replicated in a clinical BPD sample.

Third, the accuracy of transparency estimation may vary depending on participants' particular attachment style. Oxytocin, the key neuromodulator of the attachment system, has been suggested to enhance attention to social cues (Shamay-Tsoory & Abu-Akel, 2016), and activation of the attachment system may thus also enhance the salience of the "gaze" of others. In the context of transparency estimation, anxiously attached individuals' tendency to hyperactivate the attachment system may enhance the salience of the "gaze of the other" and may thus lead one to overestimate others' insight into one's inner mental state, whereas avoidant individuals who tend to deactivate the attachment system may disregard others' view and thus underestimate the observability of their mental states. Indeed, in a community sample, attachment anxiety was shown to be positively related to increased transparency overestimation (Vorauer et al., 2003), as well as to increased self–other merging (Vorauer & Cameron, 2002). Importantly, combinations of anxious and avoidant attachment are commonly found in BPD patients (Agrawal et al., 2004; Choi-Kain et al., 2009), reflecting a desire for—but a simultaneous discomfort with—closeness and attachment relationships

(Bender & Skodol, 2007). So, while BPD patients with higher attachment anxiety are expected to overestimate the transparency of their emotions, attachment avoidance may relate to transparency underestimation in BPD.

Fourth, a lack of responsiveness from important attachment figures to one's inner mental states (Biro et al., 2017; Isabella et al., 1989) may influence expectations about the transparency of those mental states. "Good enough" caregiving is assumed to be characterized by providing contingent and marked (embodied) responses to children's (facial) emotional expressions (Beebe et al., 2016; Fonagy et al., 2002). The repeated experience of having one's emotional expressions met with a contingent response from a caregiver may foster awareness about the observability of one's inner mental states by others. However, the lack of parental responsiveness experienced in a context of emotional or physical neglect may lead one to expect that others cannot "see" what one feels, possibly leading to transparency underestimation. Therefore, we expect that the pervasive attachment insecurity (Agrawal et al., 2004) and experiences of childhood trauma (Porter et al., 2020) frequently reported in individuals with BPD will relate to impairments in transparency estimation in BPD.

The present study used an adaptation of the experimental design of Barr and Kleck (1995), in replication of De Meulemeester et al. (2021). Specifically, participants' estimations of the transparency of their own emotions (estimated transparency; ET) while watching emotionally evocative movie clips were contrasted with their objective transparency (OT), operationalized as the intensity of their facial emotional expressions as analyzed using facial coding software (Noldus FaceReader). Based on the evidence and arguments presented above, we hypothesized that:

1. BPD patients would make more transparency estimation errors than HCs in terms of both overestimation (ET > OT) and underestimation (ET < OT) of transparency. This would be expressed as a larger discrepancy and less correspondence between ET and

OT in the BPD patients compared with HC participants.

2. Transparency overestimation would relate to higher SAR during the task, increased self-reported attachment anxiety, and certainty about mental states, whereas transparency underestimation would relate to low SAR, increased attachment avoidance, and uncertainty about mental states.

Method

Participants and Procedure

Participants in the BPD group were consecutively admitted patients from the University Psychiatric Hospital UPC KU Leuven, Kortenberg (Belgium) and the regional psychiatric hospital Asster in Sint-Truiden (Belgium) who met the following inclusion criteria: (a) a primary diagnosis of BPD according to the Structured Clinical Interview for DSM-5 Personality Disorders, (b) age between 18 and 60 years, and (c) speaking Dutch. In total, 86 patients were screened for BPD; 37 (43%) of these patients met the inclusion criteria and 35 (40.7%) agreed to participate in the study. Of the participants with BPD, 30 (85.7%) were female, and five (14.3%) were male. Their mean age was 27.97 years (*SD* = 8.5). The HC participants were recruited from the community via flyers and online advertisements. Inclusion criteria for the HC participants were: (a) age and sex individually matched to BPD participants and (b) speaking Dutch. The exclusion criterion was meeting the criteria for at least one personality disorder, as assessed using the Assessment of DSM-IV Personality Disorders (ADP-IV). Forty eligible participants were asked to participate, of whom 35 (87.5%) agreed and were enrolled in the study. Thirty HC participants (86.7%) were female and five (14.3%) were male, and their mean age was 27.11 years (*SD* = 8.28).

At the time of the study, only one HC participant used psychotropic medication, whereas 33 BPD participants used some type of medication. Specifically, 74.3% of the BPD participants were taking antidepressant medication, 25.7% antipsychotic medication, 11.4%

were on mood stabilizers, and 5.7% reported use of anxiolytics. Comorbidity with psychiatric symptom disorders (most commonly mood and anxiety disorders) and with other personality disorders (especially avoidant personality disorder and paranoid personality disorder) was high in the BPD patients (see Table S2 in the Supplemental Materials).

The BPD participants filled out questionnaires (see below) in a first session at the treatment facility, and completed the transparency estimation task in a second session scheduled within the same week. The HC participants filled out the questionnaires at home and were invited for a 1-hour session at the University of Leuven. Written informed consent was obtained and the study was approved by the Ethics Committee Research UZ/KU Leuven (S61144).

Measures

Transparency estimation task. This task has been described in more detail in De Meulemeester et al. (2021). Participants watched five 2-minute long film clips (see Table S1 in the Supplemental Materials for a description of the clips), which were selected scenes from feature films with attachment-related themes that are likely to elicit a strong emotional experience (Hewig et al., 2005). After each film clip, participants rated their SAR on a 7-point Likert scale ("*How strong were the emotions you experienced while watching the clip?*"). They also estimated the extent to which their emotional experience would be transparent to others using four items answered on a 10-point Likert scale (e.g., "*I think it must have been visible to others what I was feeling*", "*I think others could read on my face what I was feeling*").

Participants' facial expressions were filmed while they were watching the clips, and the intensity of their facial emotion expressions (i.e., their OT) was analyzed using automated facial coding software (FaceReader 7.1, Noldus). FaceReader analyzes facial expressions at a rate of 15 frames per second using an artificial neural network, yielding intensity scores per

frame for seven emotions (happy, sad, angry, surprised, scared, disgusted, and contempt) and a neutral expression score between 0 and 1, with 0 representing 0% agreement between the frame and the database images and 1 representing a 100% match. The facial expression scores were calibrated for each participant's own neutral facial expression to start from an individualized baseline per person. For each emotional expression (excluding the neutral expressions), an average intensity score across all frames of the clip was calculated, and the sum score of the seven different emotion categories per participant per clip was used as an overall index of the intensity of facial expressiveness. This index was used as a measure of OT per participant per clip. Generally, FaceReader has been found to achieve levels of accuracy for emotion classification of between 80% and 88% (Lewinski et al., 2014; Skiendziel et al., 2019). The image quality score (between 0 and 1) generated by FaceReader was excellent for the data in the present study (M = 0.90, SD = 0.07 in the HC group; M = 0.92, SD = 0.04 in the BPD group).

Self-report measures. The Psychiatric Diagnostic Screening Questionnaire was used to screen for 13 common psychiatric symptom disorders, with excellent reliability in the present study (126 items, Cronbach's α = 0.98) (Zimmerman & Mattia, 2001), and the ADP-IV was used to screen for personality disorders (94 items, α = 0.98) (Schotte et al., 2004). Experiences of childhood maltreatment were retrospectively assessed using the Childhood Trauma Questionnaire, of which we used the total score (25 items, α = 0.95) (Bernstein et al., 2003). Attachment-related anxiety (18 items, α = 0.96) and avoidance (18 items, α = 0.95) in intimate relationships in general were measured using the Experiences in Close Relationships-Revised questionnaire, which also achieved excellent internal consistency (Fraley et al., 2000). We measured identity diffusion using the Identity Diffusion subscale (21 items, α = 0.93) from the Inventory of Personality Organization, based on Kernberg's theory of personality organization (Lenzenweger et al., 2001). The Reflective Functioning

Questionnaire (RFQ) measured both certainty (6 items, $\alpha = 0.77$) and uncertainty (6 items, $\alpha = 0.84$) about mental states (Fonagy et al., 2016). Finally, interpersonal problems were assessed with the Inventory of Interpersonal Problems (32 items, $\alpha = 0.93$), of which we used the total score reflecting the general severity of interpersonal problems (Horowitz et al., 1988).

Data Analysis

Demographic and clinical features were analyzed using IBM SPSS Statistics for Windows, version 25 (IBM Corp., Armonk, NY, USA). Group differences were investigated with independent-samples *t*-tests for continuous variables, and chi-square tests for categorical variables; effect sizes for these analyses are expressed in terms of Cohen's *d* and Cramer's *V*, respectively. The BPD and HC groups differed significantly in educational level, but as this did not relate to any of the experimental task variables, we did not control the main analyses for this covariate. Associations between self-report questionnaire measures and experimental task variables were analyzed using Pearson correlations.

Multilevel linear models (MLMs) were used to analyze the data from the transparency estimation task with the *lme4* package in R version 3.6 (Bates et al., 2015). The analyses were controlled for differences between males and females by adding Sex as a fixed predictor in all the models. We analyzed group differences in ET, OT, and SAR. Next, group differences in transparency estimation were analyzed, calculated as the discrepancy between the *z*-score of ET per clip and the *z*-score of OT per clip (standardized relative to the total sample). Higher positive scores indicate transparency overestimation ($z\text{ET} > z\text{OT}$) and negative scores indicate transparency underestimation ($z\text{ET} < z\text{OT}$). We also analyzed group differences in the absolute value of this discrepancy score, which represents the size of the transparency error irrespective of its direction. Tukey-corrected post-hoc *t*-tests were performed on the fitted models using the *emmeans* package in R to test group differences for each film clip individually. Next, to test the hypothesis that there would be less correspondence between OT

and ET in the BPD group compared with the HC group, an MLM was fitted predicting OT based on ET and BPD group status. The time-varying predictor ET was disaggregated in a within-person centered (i.e., the amount by which a participant deviated from their own average for each clip) and a between-person centered (i.e., the amount by which a participant differed from others in the sample) part to assess both sources of information separately (Curran & Bauer, 2011). All models contained a random intercept per participant, as repeated measures were nested within individuals.

One BPD patient refused to continue the task after the first film clip, and eight observations could not be used due to technical issues with the face recordings, resulting in 12 missing values out (3.42%). In total, 338 observations from 69 unique individuals were included in the analyses. The maximum likelihood estimation method for mixed models we used can effectively deal with missing data in repeated measures (Molenberghs & Kenward, 2007). The sample size was based on an a priori power analysis (n = 68 needed to detect a between-group difference (medium size, f = 0.25) in a mixed ANOVA with five repeated measures and two groups with a power of 0.80, as calculated using Gpower 3.1 (Faul et al., 2007).

Results

Sample Description

The BPD and HC groups were matched in terms of age and sex, but the HC participants had significantly higher levels of education (see Table 1). BPD patients also had significantly higher scores than HCs on identity diffusion, interpersonal problems, childhood trauma, uncertainty about mental states, and attachment insecurity, with large effect sizes (d = 1.61–2.43) (Table 1). The HCs scored higher than the BPD patients on certainty about mental states.

Transparency Estimation

A visual summary of the distribution of the experimental task variables per group is provided in Figure 1. A detailed depiction of the facial expressions elicited by the film clips in the two groups as detected using the FaceReader software can be viewed in the Supplemental Materials (Figure S1).

Across film clips, BPD participants estimated their transparency to be significantly lower than HC participants (b = –0.93, 95% CI [–1.7; –0.18], p = 0.016). When performing Tukey-adjusted post-hoc tests per film clip, this group difference was significant for the anger-eliciting Film 2 ($t(178) = 2.48$, p = 0.014) and the happiness-eliciting Film 5 ($t(181)$ = 2.50, p = 0.013), but not for the other films (see Figure 1a). BPD participants also tended to report less SAR than HC participants, but this was not significant across all clips (b = –0.31, 95% CI [–0.801; 0.17], p = 0.20), and was significant only in response to the happiness-eliciting Film 5, which was shown last ($t(223) = 2.30$, p = 0.022; see Figure 1b). Both groups, however, did not differ in the OT of their emotional states (b = 0.01, 95% CI [–0.06: 0.08], p = 0.85), either across the film clips or for any of the individual film clips (see Figure 1c).

The discrepancy between standardized ET and OT scores per clip was calculated, with higher positive scores indicating transparency overestimation (ET > OT), and more negative scores indicating transparency underestimation (ET < OT). Across clips, BPD patients underestimated their transparency (see Figure 1d; discrepancy score <0), whereas HC participants tended to overestimate their transparency (see Figure 1d; discrepancy score >0), and this group difference almost reached statistical significance (b = –0.46, 95% CI [–0.93; 0.01], p = 0.058). There was an interaction of Group and Film showing that, as the experiment progressed, HCs increasingly overestimated their transparency, whereas BPD patients increasingly underestimated their transparency over time (b = –0.11, 95% CI [–0.22; 0.00], p = 0.049) (see Figure 1d). As a result, the group difference in transparency estimation was

significant for the positive Film 5, which was always presented last ($t(143) = 2.44$, $p = 0.016$), but not for the other individual film clips.

Contrary to our hypotheses, BPD patients did not show more transparency errors overall, calculated as the absolute value of the discrepancy between ET and OT ($b = 0.07$, 95% CI [–0.20; 0.34], $p = 0.59$). There was also no interaction of Film and Group ($b = 0.05$, 95% CI [–0.03; 0.14], $p = 0.19$). Similarly, HC participants' ET scores did not better predict OT than those of the BPD patients. In an MLM predicting OT based on ET and Group, between-subject variance in ET scores did not significantly predict OT ($b = 0.02$; 95% CI [–0.01; 0.06], $p = 0.21$), but within-subject changes in ET did ($b = 0.02$; 95% CI [0.01; 0.03], $p = 0.002$), showing that participants adjusted their ET scores over time according to changes in their actual OT. BPD group status, however, did not moderate the association between ET and OT either on the within-subject level ($b = –0.00$; 95% CI [–0.01; 0.02], $p = 0.78$) or on the between-subject level ($b = –0.01$; 95% CI [–0.06; 0.03], $p = 0.55$). BPD patients thus did not show less correspondence between ET and OT than HCs. There was no interaction with Film. This shows that the overall accuracy of transparency estimation (that is, ET scores correctly predicting OT) did not differ between the groups. Taken together, these results show that the BPD patients did not make more transparency errors than the HCs, but they showed a different type of transparency error, that is, they underestimated their transparency while HCs overestimated their transparency.

Associations Between Transparency Estimation and Self-Report Measures

The associations between the experimental task variables and the self-report measures are presented in Table 2. Within the BPD group, attachment anxiety correlated with increased SAR ($r = 0.56$, $p < 0.001$) and SAR correlated with transparency overestimation ($r = 0.40$, $p = 0.018$). In the HC group, ET was correlated negatively with attachment avoidance ($r = –0.38$,

$p = 0.02$), identity diffusion ($r = –0.51$, $p = 0.002$), and interpersonal problems ($r = –0.42$, $p = 0.012$), and positively with certainty about mental states ($r = 0.41$, $p = 0.014$).

Discussion

The present study investigated transparency estimation, that is, the ability to estimate the extent to which one's mental states are observable to others, in clinically diagnosed BPD patients and HC participants. Participants' estimations of the transparency of their emotional experience were contrasted with the OT of their emotions based on the intensity of their facial emotional expressions. Contrary to previous findings of both over- and underestimation errors in community young adults with elevated BPD features (De Meulemeester et al., 2021), the clinically diagnosed BPD patients did not show a larger discrepancy between ET and OT compared with the HCs, and thus did not exhibit more errors in transparency estimation overall. Instead, the BPD patients were found to underestimate their transparency (ET < OT), whereas the HCs overestimated their transparency (ET > OT), the latter observation being in line with previous findings of transparency overestimation in the general population (Thomas Gilovich et al., 1998; Vorauer et al., 1999; Vorauer & Cameron, 2002). Our findings of reduced feelings of transparency in BPD are in contrast with findings from other mental disorders, such as social anxiety (Brown & Stopa, 2007; Savitsky & Gilovich, 2003) and narcissism (Renier & Toma, 2022), which were found to be related to feeling *more* transparent.

In the experimental task, five emotion-eliciting film clips were always shown in a fixed order, with the four films targeting negative emotions (anger and sadness) being presented first and the film targeting a positive emotion (happiness) presented last. The group difference in transparency estimation was significant only for the positive clip, and a linear trend toward increasing transparency underestimation over time was present in the BPD patients. One possible explanation is that emotional arousal accumulated across the task,

leading to a breakdown of mentalizing (Nolte et al., 2013) and thus to worse transparency estimation. However, the self-reported emotional arousal (i.e., SAR) scores of the participants did not support this. Another possibility is that BPD patients may be less aware of their own positive emotion expressions than of their negative emotion expressions (Hooley & Masland, 2018). The positive film clip that was shown at the end of the task may also have been surprising and unexpected after four negative film stimuli, making correct transparency estimation more challenging.

Overall, the fact that BPD patients underestimated how observable their emotions are to others is in line with a recent study of adolescent BPD patients' lack of insight into their own level of stress (Bourvis et al., 2020). BPD patients are often found to report low subjective emotions while displaying high objective emotional arousal (for a review, see Perez-Rodriguez et al. (2018)). Recently, research has focused on BPD patients' problems in detecting their own internal bodily signals, called interoception (Back & Bertsch, 2020; Löffler et al., 2018). The finding of reduced awareness of their own facial emotion expressions in the BPD patients is also in line with recent conceptualizations of BPD as being characterized by impairments in embodied mentalizing (Fotopoulou & Tsakiris, 2017; Neustadter et al., 2019). The present study contributes to this body of research on low self-awareness in BPD by showing that BPD patients assume their emotions are not observable to others, while their emotions are in fact being communicated through their facial expressions. Two possible mechanisms explaining transparency underestimation in BPD emerged from the findings of this study.

First, prior experiences of their emotional needs being ignored or overlooked by others may lead BPD patients to expect that others will not "see" how they are feeling. Indeed, BPD patients estimated their transparency to be significantly lower than the HCs, although they were not objectively less transparent. In the HC group, estimating one's emotions to be less

transparent was associated with attachment avoidance, and in the total sample, it was also associated with childhood emotional neglect. These associations were not present in the BPD group, which we did not expect. Nevertheless, compared with HCs, BPD patients had significantly higher levels of attachment avoidance and emotional neglect, with large effect sizes (d = 1.69 and 1.82, respectively). Avoidance of or discomfort with attachment is prevalent in BPD (Agrawal et al., 2004; Levy et al., 2005) and entails a defensive deactivation of attachment to protect oneself from being disappointed by others, implying an expectation that others will not be responsive to one's needs. Emotional neglect was found to be highly predictive of BPD in several studies (Porter et al., 2020; Zanarini et al., 2000), as were other features relating to an invalidating social environment (Crowell et al., 2009; Musser et al., 2018). Prior experiences of one's emotional experiences being overlooked by significant others may lead one to feel "invisible" to others. However, these interpretations are speculative, since the direction of causality cannot be determined from these associations.

Second, BPD patients reported lower levels of SAR compared with HCs, on a trend level, suggesting that their own emotions were perhaps not very salient to them. As hypothesized, reporting higher subjective emotional arousal during the task was related to increased transparency overestimation in the BPD group, in line with earlier findings (De Meulemeester et al., 2021; Vorauer et al., 1999). BPD patients' low awareness of their own emotional state may erroneously lead them to assume that their emotions are not observable to others. However, in the BPD group, attachment anxiety was correlated with reports of higher SAR, which in turn related to transparency overestimation. This suggests a potential (causal) pathway among the BPD sample: high levels of attachment anxiety may lead to higher experienced emotional intensity in this task, which may lead to the overestimation of the transparency of these emotions. There were large amounts of variation in transparency estimation in the BPD group (see Figure 1), and this may be in part explained by differences

in attachment style and SAR between BPD patients. Indeed, previous studies in community adults found that attachment anxiety related to transparency overestimation (Vorauer et al., 2003) and to higher emotional awareness and better interoceptive abilities, whereas attachment avoidance related to lower emotional awareness and worse interoception (Fantini-Hauwel et al., 2012; Oldroyd et al., 2019).

There was a significant positive association between certainty about mental states and transparency overestimation within the HC group, but, contrary to our expectation, not within the BPD group. In the general population, being extremely certain about one's own mental states may lead to a failure to recognize that these mental states are opaque unless they are communicated (Bo et al., 2017; Sharp & Vanwoerden, 2015). The BPD patients in the present study scored high on uncertainty about mental states, which is assumed to reflect hypomentalizing (Fonagy et al., 2016). However, against our expectations, uncertainty about mental states did not relate to transparency estimation, neither in the BPD nor in the HC group. Also, negative associations between identity and interpersonal problems and transparency feelings were only found in the HC group, suggesting that feeling less transparent is maladaptive and related to poor mental health outcomes in the general population. Surprisingly, increased severity of these symptoms did not relate to larger transparency errors in the BPD patients.

Nevertheless, the fact that BPD patients underestimated the transparency of their emotions compared to HCs may have a negative impact on their social interactions, which typically involve high levels of rejection and low reciprocity (Lis & Bohus, 2013). Indeed, emotions that are expressed by BPD patients through nonverbal signals outside their awareness may influence their interaction partners' emotional state (Dimberg et al., 2000; Prochazkova & Kret, 2017). If the BPD patient does not "own up" to their own emotional experience, the interaction partner may confuse these emotions as their own (a phenomenon

called emotional contagion (Hatfield et al., 1993)). As such, transparency underestimation may be linked to projective identification (Gallese, 2009; Zanarini et al., 2013). The influx of confusing or negative emotions experienced when interacting with BPD patients may lead others to distance themselves and ultimately reject individuals with BPD (Hepp et al., 2019; Hepp et al., 2017). If BPD patients are unaware of the emotions they express, they may be less able to modulate what they communicate to others and less able to understand others' reactions to them (Qu et al., 2017). The experience of problematic interactions with others may further exacerbate their expectation that others do not "see" or understand how they feel. Future research should investigate the impact of transparency estimation errors on interaction partners' emotions and on the quality of interactions. Also, investigating transparency estimation using personalized video stimuli or during live social interactions would enhance the ecological validity of the paradigm.

Despite its strengths, there are several limitations to the present study. First, although the BPD and HC participants were matched on age and sex, the HC participants had significantly higher levels of education. However, level of education was not correlated with any of the experimental variables. Most BPD patients used some type of psychotropic medication, compared with only one of the HCs. Medication use is not uncommon in patients involved in inpatient treatment, but this difference may affect the internal validity of the findings and hamper their generalizability to nonmedicated BPD patients. Replication in nonmedicated BPD participants and carefully matched control groups is necessary. Second, the participants were predominantly female (82.9%). We controlled all our analyses for differences in sex, however, future research should aim to achieve a more balanced distribution so that differences between men and women with BPD can be systematically studied. Third, the HCs were screened for psychopathology via self-report measures instead of interview-based measures and the study did not include a clinical control sample. Fourth,

this study focused only on facial expressions and did not consider other determinants of objective transparency, such as head pose, blushing, or vocalizations. Fifth, part of the experiment relied on self-report, which may be biased, particularly in BPD patients. Indeed, a recent study found a small correlation between BPD features and extremity responding (Kaufman et al., 2022). Finally, the transparency estimation task required participants to estimate the transparency of their emotional state from the perspective of an abstract, imagined "other". Future research should consider investigating transparency estimation in BPD during actual social interactions.

In conclusion, this study shows that BPD patients underestimate the transparency of their emotional mental states, which may stem from prior experiences in which their emotions and needs were invalidated or overlooked. As a result, BPD patients may believe their emotions are not transparent, while they may in fact be observable to others, at least based on their facial emotion expressions. This reduced ability to represent their own emotions from the perspective of others may reflect problems with embodied mentalizing and self-awareness in BPD, and may negatively impact their social interactions. The present study shows that BPD may be associated with problems in representing the "self" from the perspective of others, and specifically with a lack of awareness of the intensity of their own facial emotion expressions.

References

Agrawal, H. R., Gunderson, J., Holmes, B. M., & Lyons-Ruth, K. (2004). Attachment studies with borderline patients: A review. *Harvard Review of Psychiatry, 12*(2), 94-104. https://doi.org/10.1080/10673220490447218

Back, S. N., & Bertsch, K. (2020). Interoceptive Processing in Borderline Personality Pathology: a Review on Neurophysiological Mechanisms. *Current Behavioral Neuroscience Reports*(7), 232-238. https://doi.org/10.1007/s40473-020-00217-2

Barr, C. L., & Kleck, R. E. (1995). Self-other perception of the intensity of facial expressions of emotion: do we know what we show? *Journal of Personality and Social Psychology*, *68*(4), 608-618. https://doi.org/10.1037/0022-3514.68.4.608

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A systems view of mother-infant face-to-face communication. *Developmental Psychology*, *52*(4), 556-571. https://doi.org/10.1037/a0040085

Bender, D. S., & Skodol, A. E. (2007). Borderline personality as a self-other representational disturbance. *Journal of Personality Disorders, 21*, 500-517. https://doi.org/10.1521/pedi.2007.21.5.500

Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J., Handelsman, L., Medrano, M., Desmond, D., & Zule, W. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse & Neglect*, *27*(2), 169-190. https://doi.org/10.1016/S0145-2134(02)00541-0

Biro, S., Alink, L. R. A., Huffmeijer, R., Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2017). Attachment quality is related to the synchrony of mother and infant monitoring patterns. *Attachment & Human Development*, *19*(3), 243-258. https://doi.org/10.1080/14616734.2017.1302487

Bo, S., Sharp, C., Fonagy, P., & Kongerslev, M. (2017). Hypermentalizing, attachment, and epistemic trust in adolescent BPD: Clinical illustrations. *Personality disorders, 8*, 172-182. https://doi.org/10.1037/per0000161

Bourvis, N., Aouidad, A., Spodenkiewicz, M., Palestra, G., Aigrain, J., Baptista, A., Benoliel, J.-J., Chetouani, M., & Cohen, D. (2020). Adolescents with borderline personality

disorder show a higher response to stress but a lack of self-perception: Evidence through affective computing. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 110095. https://doi.org/https://doi.org/10.1016/j.pnpbp.2020.110095

Brown, M. A., & Stopa, L. (2007). The spotlight effect and the illusion of transparency in social anxiety. *J Anxiety Disord, 21*(6), 804-819. https://doi.org/10.1016/j.janxdis.2006.11.006

Cameron, J. J., & Vorauer, J. D. (2008). Feeling Transparent: On Metaperceptions and Miscommunications. *Social and Personality Psychology Compass, 2*(2), 1093-1108. https://doi.org/10.1111/j.1751-9004.2008.00096.x

Choi-Kain, W. L., Fitzmaurice, M. G., Zanarini, C. M., Laverdière, G. O., & Gunderson, G. J. (2009). The Relationship Between Self-Reported Attachment Styles, Interpersonal Dysfunction, and Borderline Personality Disorder. *Journal of Nervous and Mental Disease, 197*(11), 816-821. https://doi.org/10.1097/NMD.0b013e3181bea56e

Colle, L., Gabbatore, I., Riberi, E., Borroz, E., Bosco, F. M., & Keller, R. (2018). Mindreading abilities and borderline personality disorder: A comprehensive assessment using the Theory of Mind Assessment Scale. *Psychiatry Research, 272*, 609-617. https://doi.org/10.1016/j.psychres.2018.12.102

Crowell, S. E., Beauchaine, T. P., & Linehan, M. M. (2009). A biosocial developmental model of borderline personality: Elaborating and extending Linehan's theory. *Psychological bulletin, 135*(3), 495-510. https://doi.org/10.1037/a0015616

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual review of psychology, 62*, 583-619. https://doi.org/10.1146/annurev.psych.093008.100356

De Meulemeester, C., Lowyck, B., Boets, B., Van der Donck, S., & Luyten, P. (2021). Do my emotions show or not? Problems with transparency estimation in women with

borderline personality disorder features. *Personality Disorders-Theory Research and Treatment*, Online first publication ahead of print. https://doi.org/10.1037/per0000504

Derks, Y., Westerhof, G. J., & Bohlmeijer, E. T. (2017). A Meta-analysis on the Association Between Emotional Awareness and Borderline Personality Pathology. *Journal of Personality Disorders, 31*(3), 362-384. https://doi.org/10.1521/pedi_2016_30_257

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*(1), 86-89. https://doi.org/10.1111/1467-9280.00221

Fantini-Hauwel, C., Boudoukha, A. H., & Arciszewski, T. (2012). Adult attachment and emotional awareness impairment: a multimethod assessment. *Socioaffective Neuroscience & Psychology, 2*(1), 10744. https://doi.org/10.3402/snp.v2i0.10744

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/bf03193146

Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2002). *Affect Regulation, Mentalization, and the Development of the Self*. Routledge.

Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology, 21*(4), 1355-1381. https://doi.org/10.1017/S0954579409990198

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y., Warren, F., Howard, S., Ghinai, R., Fearon, P., & Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. *PLoS One, 11*. https://doi.org/10.1371/journal.pone.0158678

Fotopoulou, A., & Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference. *Neuropsychoanalysis, 19*(1), 3-28. https://doi.org/10.1080/15294145.2017.1294031

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology, 78*, 350-365. https://doi.org/http://dx.doi.org/10.1037/0022-3514.78.2.350

Gallese, V. (2009). Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification. *Psychoanalytic Dialogues, 19*(5), 519-536. https://doi.org/10.1080/10481880903231910

Gilovich, T., Medvec, V. H., & Savitsky, K. (1998). The Illusion of Transparency: Biased Assessments of Others' Ability to Read One's Emotional States. *Journal of Personality and Social Psychology, 75*, 332-346. https://doi.org/10.1037/0022-3514.75.2.332

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: biased assessments of others' ability to read one's emotional states [Research Support, U.S. Gov't, Non-P.H.S.

Research Support, U.S. Gov't, P.H.S.]. *Journal of Personality and Social Psychology, 75*(2), 332-346. http://www.ncbi.nlm.nih.gov/pubmed/9731312

Grant, B. F., Chou, S. P., Goldstein, R. B., Huang, B., Stinson, F. S., Saha, T. D., Smith, S. M., Dawson, D. A., Pulay, A. J., Pickering, R. P., & Ruan, W. J. (2008). Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry, 69*(4), 533-545. https://doi.org/10.4088/jcp.v69n0404

Gunderson, J. G., Herpertz, S. C., Skodol, A. E., Torgersen, S., & Zanarini, M. C. (2018). Borderline personality disorder. *Nature Reviews Disease Primers, 4*(1), 18029. https://doi.org/10.1038/nrdp.2018.29

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional Contagion. *Current Directions in Psychological Science, 2*(3), 96-99. http://www.jstor.org/stable/20182211

Hepp, J., Gebhardt, S., Kieslich, P. J., Storkel, L. M., & Niedtfeld, I. (2019). Low positive affect display mediates the association between borderline personality disorder and negative evaluations at zero acquaintance. *Borderline Personality Disorder and Emotion Dysregulation, 6*, 4. https://doi.org/10.1186/s40479-019-0103-6

Hepp, J., Lane, S. P., Carpenter, R. W., Niedtfeld, I., Brown, W. C., & Trull, T. J. (2017). Interpersonal problems and negative affect in Borderline Personality and Depressive Disorders in daily life. *Clinical psychological science : a journal of the Association for Psychological Science, 5*(3), 470-484. https://doi.org/10.1177/2167702616677312

Hewig, J., Hagemann, D., Seifert, J., Gollwitzer, M., Naumann, E., & Bartussek, D. (2005). A revised film set for the induction of basic emotions. *Cognition & Emotion, 19*(7), 1095-1109. https://doi.org/10.1080/02699930541000084

Hooley, J., & Masland, S. (2018). Positive Emotion in Borderline Personality Disorder. In.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*(6), 885-892. https://doi.org/10.1037//0022-006x.56.6.885

Isabella, R. A., Belsky, J., & von Eye, A. (1989). Origins of infant-mother attachment: An examination of interactional synchrony during the infant's first year. *Developmental Psychology, 25*(1), 12-21. https://doi.org/10.1037/0012-1649.25.1.12

Kaufman, E., Beeney, J., & Stepp, S. (2022). Self-report response style bias and borderline personality features. *Current Psychology*, 1-9. https://doi.org/10.1007/s12144-022-03122-x

Korzekwa, M. I., Dell, P. F., Links, P. S., Thabane, L., & Webb, S. P. (2008). Estimating the prevalence of borderline personality disorder in psychiatric outpatients using a two-phase procedure. *Comprehensive Psychiatry, 49*(4), 380-386. https://doi.org/10.1016/j.comppsych.2008.01.007

Lenzenweger, M. F., Clarkin, J. F., Kernberg, O. F., & Foelsch, P. A. (2001). The Inventory of Personality Organization: Psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis proneness, and self-domains in a nonclinical sample. *Psychological Assessment, 13*(4), 577-591. https://doi.org/10.1037//1040-3590.13.4.577

Levy, K., Meehan, K., Weber, M., Reynoso, J., & Clarkin, J. (2005). Attachment and Borderline Personality Disorder: Implications for Psychotherapy. *Psychopathology, 38*, 64-74. https://doi.org/10.1159/000084813

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated Facial Coding: Validation of Basic Emotions and FACS AUs in FaceReader. *Journal of Neuroscience Psychology and Economics, 7*(4), 227-236. https://doi.org/10.1037/npe0000028

Lis, S., & Bohus, M. (2013). Social interaction in borderline personality disorder. *Current Psychiatry Reports, 15*, 338. https://doi.org/10.1007/s11920-012-0338-z

Löffler, A., Foell, J., & Bekrater-Bodmann, R. (2018). Interoception and its interaction with self, other, and emotion processing: Implications for the understanding of psychosocial deficits in borderline personality disorder. *Current Psychiatry Reports, 20*(4), 28. https://doi.org/10.1007/s11920-018-0890-2

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). The Mentalizing Approach to Psychopathology: State of the Art and Future Directions. *Annual Review of Clinical Psychology*. https://doi.org/10.1146/annurev-clinpsy-071919-015355

Molenberghs, G., & Kenward, M. (2007). *Missing Data in Clinical Studies*. https://doi.org/10.1002/9780470510445

Morey, L. C. (2014). Borderline features are associated with inaccurate trait self-estimations. *Borderline Personal Disord Emot Dysregul, 1*, 4. https://doi.org/10.1186/2051-6673-1-4

Musser, N., Zalewski, M., Stepp, S., & Lewis, J. (2018). A systematic review of negative parenting practices predicting borderline personality disorder: Are we measuring biosocial theory's 'invalidating environment'? *Clinical Psychology Review, 65*, 1-16. https://doi.org/10.1016/j.cpr.2018.06.003

Németh, N., Mátrai, P., Hegyi, P., Czéh, B., Czopf, L., Hussain, A., Pammer, J., Szabó, I., Solymár, M., Kiss, L., Hartmann, P., Szilágyi, Á. L., Kiss, Z., & Simon, M. (2018). Theory of mind disturbances in borderline personality disorder: A meta-analysis. *Psychiatry Research, 270*, 143-153. https://doi.org/https://doi.org/10.1016/j.psychres.2018.08.049

Neustadter, E. S., Fotopoulou, A., Steinfeld, M., & Fineberg, S. K. (2019). Mentalization and embodied selfhood in Borderline Personality Disorder. *PsyArXiv Preprint*. https://doi.org/10.31234/osf.io/9z3nq

New, A. S., aan het Rot, M., Ripoll, L. H., Perez-Rodriguez, M. M., Lazarus, S., Zipursky, E., Weinstein, S. R., Koenigsberg, H. W., Hazlett, E. A., Goodman, M., & Siever, L. J. (2012). Empathy and alexithymia in borderline personality disorder: clinical and laboratory measures. *Journal of Personality Disorders, 26*(5), 660-675. https://doi.org/10.1521/pedi.2012.26.5.660

Nolte, T., Bolling, D. Z., Hudac, C. M., Fonagy, P., Mayes, L., & Pelphrey, K. A. (2013). Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Frontiers in Human Neuroscience, 7*, 816-816. https://doi.org/10.3389/fnhum.2013.00816

Oldroyd, K., Pasupathi, M., & Wainryb, C. (2019). Social Antecedents to the Development of Interoception: Attachment Related Processes Are Associated With Interoception. *Frontiers in Psychology, 10*, 712-712. https://doi.org/10.3389/fpsyg.2019.00712

Perez-Rodriguez, M. M., Bulbena-Cabre, A., Bassir Nia, A., Zipursky, G., Goodman, M., & New, A. S. (2018). The Neurobiology of Borderline Personality Disorder [Review]. *Psychiatr Clin North Am, 41*(4), 633-650. https://doi.org/10.1016/j.psc.2018.07.012

Petersen, R., Brakoulias, V., & Langdon, R. (2016). An experimental investigation of mentalization ability in borderline personality disorder. *Comprehensive Psychiatry, 64*, 12-21. https://doi.org/10.1016/j.comppsych.2015.10.004

Porter, C., Palmier-Claus, J., Branitsky, A., Mansell, W., Warwick, H., & Varese, F. (2020). Childhood adversity and borderline personality disorder: a meta-analysis. *Acta psychiatrica Scandinavica, 141*(1), 6-20. https://doi.org/10.1111/acps.13118

Prochazkova, E., & Kret, M. E. (2017). Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience and Biobehavioral Reviews, 80*, 99-114. https://doi.org/https://doi.org/10.1016/j.neubiorev.2017.05.013

Qu, F., Yan, W.-J., Chen, Y.-H., Li, K., Zhang, H., & Fu, X. (2017). "You Should Have Seen the Look on Your Face…": Self-awareness of Facial Expressions. *Frontiers in Psychology, 8*, 832-832. https://doi.org/10.3389/fpsyg.2017.00832

Renier, L. A., & Toma, C. (2022). I am a great open book: Narcissistic individuals feel transparent toward others. *Personality and Individual Differences, 192*, 111585. https://doi.org/https://doi.org/10.1016/j.paid.2022.111585

Savitsky, K., & Gilovich, T. (2003). The illusion of transparency and the alleviation of speech anxiety. *Journal of Experimental Social Psychology, 39*, 618-625. https://doi.org/10.1016/s0022-1031(03)00056-8

Schotte, C., Doncker, D., Dmitruk, D., Van Mulders, I., D'Haenen, H., & Cosyns, P. (2004). The ADP-IV questionnaire: Differential validity and concordance with the semi-structured interview. *Journal of Personality Disorders, 18*, 405-419. https://doi.org/10.1521/pedi.18.4.405.40348

Shamay-Tsoory, S. G., & Abu-Akel, A. (2016). The Social Salience Hypothesis of Oxytocin. *Biological Psychiatry, 79*(3), 194-202. https://doi.org/10.1016/j.biopsych.2015.07.020

Sharp, C., & Vanwoerden, S. (2015). Hypermentalizing in Borderline Personality Disorder: A Model and Data. *Journal of Infant, Child, and Adolescent Psychotherapy, 14*, 33-45. https://doi.org/10.1080/15289168.2015.1004890

Skiendziel, T., Rösch, A. G., & Schultheiss, O. C. (2019). Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring. *PLoS One, 14*(10), e0223905. https://doi.org/10.1371/journal.pone.0223905

Van Boven, L., Gilovich, T., & Medvec, V. (2003). The Illusion of Transparency in Negotiations. *Negotiation Journal, 19*, 117-131. https://doi.org/10.1023/A:1023693700484

Vorauer, J., Ross, M., Bailis, D., Baldwin, M., Buehler, R., Mcfarland, C., & Savitsky, K. (1999). Self-Awareness and Feeling Transparent: Failing to Suppress One's Self.

Journal of Experimental Social Psychology, *35*, 415-440. https://doi.org/10.1006/jesp.1999.1388

Vorauer, J. D., & Cameron, J. J. (2002). So close, and yet so far: does collectivism foster transparency overestimation? *Journal of Personality and Social Psychology*, *83*(6), 1344-1352. http://www.ncbi.nlm.nih.gov/pubmed/12500816

Vorauer, J. D., Cameron, J. J., Holmes, J. G., & Pearce, D. G. (2003). Invisible overtures: fears of rejection and the signal amplification bias. *Journal of Personality and Social Psychology*, *84*(4), 793-812. https://doi.org/10.1037/0022-3514.84.4.793

Zanarini, M. C., Frankenburg, F. R., & Fitzmaurice, G. (2013). Defense mechanisms reported by patients with borderline personality disorder and axis II comparison subjects over 16 years of prospective follow-up: description and prediction of recovery. *The American Journal of Psychiatry*, *170*(1), 111-120. https://doi.org/10.1176/appi.ajp.2012.12020173

Zanarini, M. C., Frankenburg, F. R., Reich, D. B., Marino, M. F., Lewis, R. E., Williams, A. A., & Khera, G. S. (2000). Biparental failure in the childhood experiences of borderline patients. *Journal of Personality Disorders*, *14*(3), 264-273. https://doi.org/10.1521/pedi.2000.14.3.264

Zimmerman, M., & Mattia, J. I. (2001). A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. *Archives of General Psychiatry*, *58*, 787-794. https://doi.org/10.1001/archpsyc.58.8.787

Table 1. Demographic and Clinical Features of BPD Patients and Healthy Controls

	BPD (*n* = 35/31)	**HC (*n* = 35)**	**Comparison**	
	Mean (*SD*) / Count (%)	Mean (SD) / Count (%)	Test statistic	Cohen's *d* / Cramer's *V*
Female	30 (85.71%)	30 (85.71%)		
Age	27.97 (8.5)	27.11 (8.28)	$t(68) = 0.427$, $p = 0.67$	$d = 0.1$
Education level			$\chi^2(3) = -19.3$, $p < 0.001$	$V = 0.52$
No high school	5 (14.3%)	1 (2.9%)		
Only high school	18 (51.4%)	4 (11.4%)		
Higher education	4 (11.4%)	10 (28.6%)		
University	8 (22.9%)	20 (57.6%)		
Medication	33 (97.1%)	1 (2.9%)	$\chi^2(1) = -58.6$, $p < 0.001$	$V = 0.91$
Attachment anxiety	5.35 (1.16)	3.18 (1.26)	$t(64) = 7.221$, $p < 0.001$	$d = 1.78$
Attachment avoidance	5.03 (0.85)	3.18 (1.26)	$t(64) = 6.872$, $p < 0.001$	$d = 1.69$
Childhood trauma	66.58 (18.67)	39.34 (15.27)	$t(64) = 6.515$, $p < 0.001$	$d = 1.61$
Emotional abuse	17.58 (5.03)	9.34 (5.04)	$t(64) = 6.63$, $p < 0.001$	$d = 1.63$
Emotional neglect	19.45 (4.09)	10.68 (5.35)	$t(64) = 7.39$, $p < 0.001$	$d = 1.82$
Certainty about MS	0.47 (0.51)	1.03 (0.73)	$t(60) = -3.64$, $p = 0.001$	$d = -0.88$
Uncertainty about MS	1.61 (0.80)	0.44 (0.38)	$t(42) = 7.64$, $p < 0.001$	$d = 1.96$
Identity diffusion	72.06 (14.48)	45.00 (9.78)	$t(64) = 9.86$, $p < 0.001$	$d = 2.43$
Interpersonal problems	63.51 (16.83)	31.57 (17.74)	$t(64) = 7.53$, $p < 0.001$	$d = 1.86$

Note. HC = healthy controls; MS = mental states.

Table 2. Pearson Correlations Between Experimental Task Variables and Self-Report Measures Within the BPD group and the HC group.

	BPD group					**HC group**				
	ET	OT	SAR	zET_min_zOT	| zET_min_zOT |	ET	OT	SAR	zET_min_zOT	| zET_min_zOT |
Subjective arousal	**0.60*** **	0.08	1	**0.40***	0.22	**0.54*** **	0.06	1	0.28	0.29
Attachment anxiety	0.23	–0.00	**0.56*** **	0.16	0.05	–0.30	–0.10	–0.06	–0.10	–0.10
Attachment avoidance	–0.06	–0.11	–0.11	0.03	–0.05	**–0.38***	0.06	–0.30	–0.29	–0.19
Total childhood trauma	0.23	–0.02	0.07	0.20	0.01	–0.00	0.05	0.28	–0.04	–0.15
Emotional abuse	0.25	0.07	0.09	0.15	0.08	–0.01	0.03	0.18	–0.02	–0.12
Emotional neglect	–0.04	–0.25	–0.12	0.14	–0.09	–0.24	0.00	0.15	–0.14	–0.17
Uncertainty about mental states	0.09	0.10	0.35	–0.14	–0.22	–0.31	–0.14	–0.14	–0.09	–0.12
Certainty about mental states	0.09	–0.03	–0.23	0.09	0.18	**0.41***	0.05	–0.01	0.20	–0.02
Identity diffusion	–0.01	0.20	0.12	–0.15	–0.19	**–0.51** **	–0.06	–0.09	–0.25	–0.12
Interpersonal problems	–0.01	0.08	0.08	–0.08	–0.06	**–0.42***	–0.10	–0.23	–0.16	–0.14

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. ET = estimated transparency; OT = objective transparency; SAR = subjective arousal; zET_min_zOT = discrepancy between standardized ET and OT; | zET_min_zOT | = absolute value of the discrepancy between standardized ET and OT.

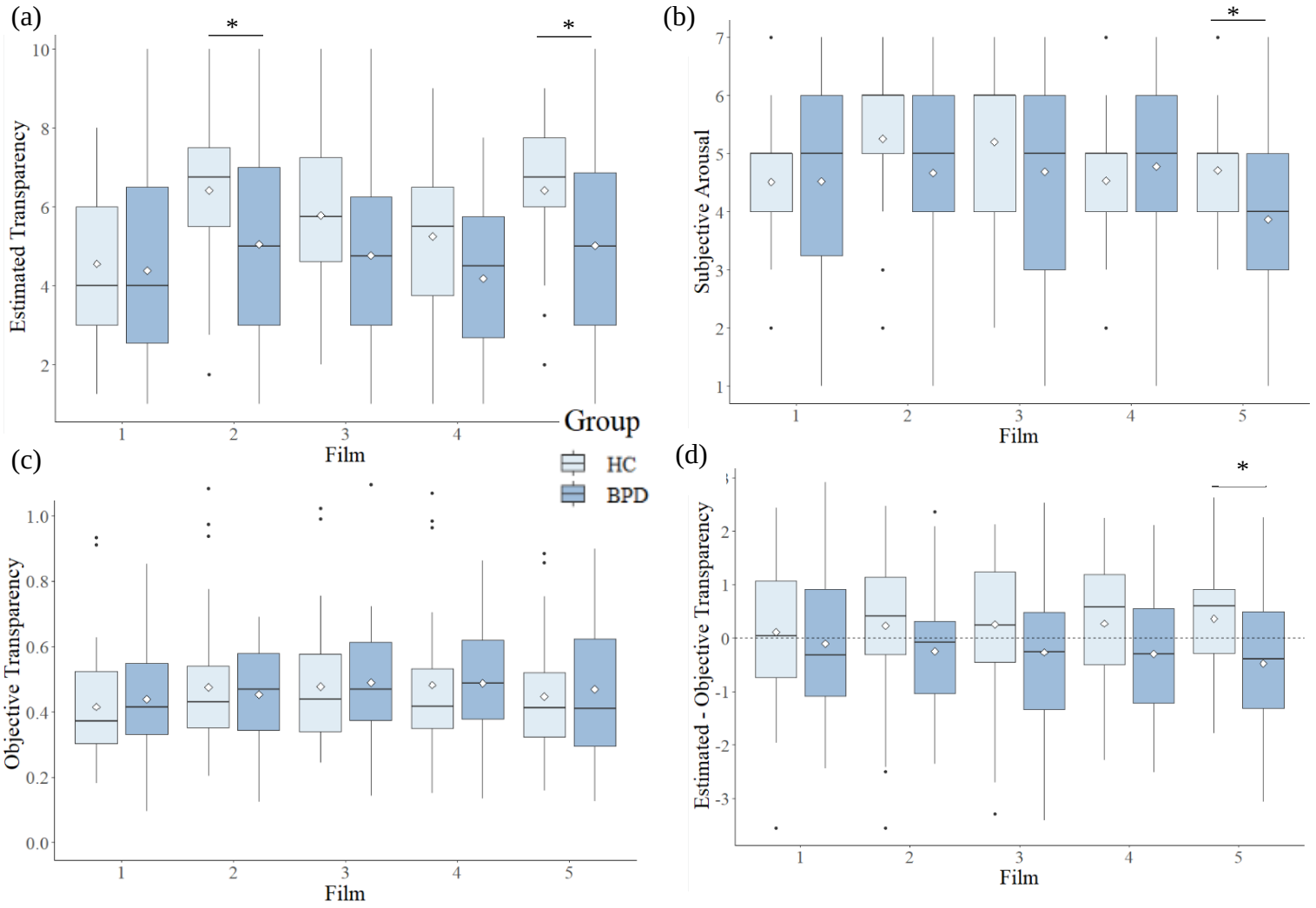

Figure 1. Boxplots of the experimental task variables (a) estimated transparency (ET), (b) subjective arousal (SAR), and (c) objective transparency (OT), and (d) the discrepancy between standardized ET and OT scores, per film clip and per group (healthy controls [HC] = light blue; BPD patients = dark blue). The five film clips targeted different emotions (1 = sadness, 2 = anger, 3 = sadness, 4 = anger, 5 = happiness) and were always shown in a fixed order. The boxes extend from the lower to the upper quartile values, with the median of each group represented by the horizontal lines. Mean values are displayed by the white diamond. Significant group differences ($p < 0.5$) are indicated with asterisks.