Do my emotions show or not? Problems with transparency estimation in women with borderline personality disorder features

Celine De Meulemeester [1*], Benedicte Lowyck [2], Bart Boets[3], Stephanie Van der Donck[3] 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, 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, London, UK

In press: Personality Disorders: Theory, Treatment and Research

Author Note

We are grateful to all our participants, and to Michelle Vanhelden and Nil Anarat for their help in the data collection.

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

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).

The findings have been presented in an oral presentation at the 2019 Conference of the International Society for the Study of Personality Disorders (ISSPD) in Vancouver (Canada).

* Corresponding author. 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

Transparency estimation, that is, estimating the extent to which one's mental states are observable to others, requires the simultaneous representation of the self and of others' perspective on the self. Individuals with borderline personality disorder (BPD) have difficulty integrating multiple perspectives when mentalizing, which may be reflected in impaired transparency estimation. Sixty-two participants high and low in BPD features watched emotionally evocative video clips, and estimated the transparency of their emotional experience while facial expression coding software (FaceReader) quantified their objective transparency. Individuals high in BPD features showed a larger discrepancy between estimated and objective transparency than individuals low in BPD features, showing that they both over- and underestimated their transparency. Indeed, estimated transparency positively predicted objective transparency in individuals low in BPD features, but not in individuals high in BPD features. Moreover, the ability to estimate intraindividual variability in one's own objective transparency was moderated by self-reported arousal in the participants high in BPD features. Impairments in transparency estimation were correlated with self-report measures of borderline features, attachment, and mentalizing. In conclusion, we found that borderline features relate to a reduced capacity to estimate the extent to which one's own emotional states are observable to others. Although replication in clinical samples of BPD patients is needed, the present study provides evidence for problems in mentalizing the (embodied) self from another person's perspective in BPD.

Keywords: transparency estimation; borderline personality disorder; facial emotion expression; mentalizing; self-other distinction; self-other differentiation

Borderline personality disorder (BPD) is a pervasive disorder characterized by interpersonal hypersensitivity, an unstable sense of self, impulsivity, and emotional dysregulation (Gunderson et al., 2018). The centrality of self- and interpersonal dysfunction in BPD has been emphasized in Criterion A of the DSM-5 Alternative Model for Personality Disorders (American Psychiatric Association, 2013), and these problems may be explained, in part, by BPD patients' profound difficulties in mentalizing, that is, reflecting on the internal mental states of self and others (Fonagy & Luyten, 2009; Luyten, De Meulemeester, et al., 2020). Specifically, hyper-reactivity in neural networks involved in sharing others' mental representations and impairments in explicit mental-state attribution has been suggested to make BPD patients vulnerable to conflating their own and others' mental states (Herpertz et al., 2018; Herpertz et al., 2014; Luyten & Fonagy, 2015; Ripoll et al., 2013). This is in line with theoretical accounts proposing impairments in self–other distinction to be a key feature of BPD (Bender & Skodol, 2007; Neustadter, Fotopoulou, et al., 2019).

Impairments in mentalizing with regard to the self and others may hamper BPD patients' capacity for transparency estimation, that is, the ability to estimate the extent to which one's own mental states are observable to others (Gilovich et al., 1998; Vorauer et al., 1999). Adequately estimating the extent to which one's mental states are transparent to others requires awareness of one's own mental state and processing of the interoceptive information associated with bodily signals (e.g., facial muscle movements) through which the mental state may be communicated to others (i.e., embodied, affective mentalizing of the "self"). At the same time, one needs to be able to take a third-person perspective on the self in order to gauge how one is perceived by others (i.e., cognitive, reflective mentalizing of the "other"). Transparency estimation thus involves a complex set of mentalizing abilities and the balanced integration of information based on these different capacities. Yet, BPD patients are typically found to have difficulty integrating different mentalizing dimensions and differentiating their

own and others' perspectives (Beeney et al., 2016; Luyten, Campbell, Allison, et al., 2020; Petersen et al., 2016).

First, mentalizing the self has been found to be altered in BPD. On the one hand, the high emotional reactivity (Scott et al., 2013) typical of BPD often results in extremely high levels of self-consciousness. Feelings of shame, for instance, are often very intense in BPD patients, making them painfully aware of any perceived shortcomings or flaws (Peters & Geiger, 2016). This enhanced salience of their self-experience may lead them to overestimate the transparency of these experiences. Indeed, in normal controls, increased self-salience was positively related to transparency overestimation, purportedly because of a failure to inhibit the highly salient self-experience in order to take the perspective of the other and to recognize that the experience is not shared (Vorauer et al., 1999). On the other hand, a reduced ability to identify and label one's own mental states, in particular emotional states (i.e., alexithymia), has also been found in individuals with BPD (Derks et al., 2017), as has reduced interoceptive awareness, that is, a reduced awareness of signals arising from within the body (Löffler et al., 2018). When individuals with BPD are not able to represent their own mental and bodily state, they may assume that these states are not transparent, when they may in fact be communicating mental states through various bodily signals (e.g., facial muscle movement, posture, muscle tone, etc.), resulting in an underestimation of their transparency.

Second, BPD also seems to be characterized by problems with perspective taking. The ability to shift one's perspective from the "self" to another person enables one to disengage from one's own private self-knowledge in order to see the self from the perspective of others (Chambers et al., 2008). The absence of this capacity may result in the egocentric assumption that one's own mental states are shared, and thus transparent, to others (i.e., transparency overestimation). Indeed, shifting attention to the perspective of the interaction partner was found to significantly reduce the overestimation of transparency (Vorauer & Sucharyna,

2013). Shifting from representing "self" to representing "other" is an effortful process that implies self–other distinction (Lamm et al., 2016a; Sowden & Shah, 2014). Self-report (Harari et al., 2010), experimental (New et al., 2012; Petersen et al., 2016), and interview-based (Colle et al., 2018) investigations have found impairments in BPD patients' capacity to take the perspective of others, which may extend to taking a third-person perspective on the (embodied) self and thus hamper their ability to estimate how they are perceived by others. This absence of third-person perspective taking on the self may lead individuals with BPD to overestimate the transparency of their mental states when their own experience is highly salient to them, and to underestimate their transparency when they are unaware of their own mental state and of the bodily signals that may communicate it to others. These transparency estimation errors may make BPD patients less able to gauge how they are perceived by others, which may result in misunderstanding and interpersonal conflict (Cameron & Vorauer, 2008).

From a developmental perspective, studies suggest that expectations about the transparency of one's own mental states may in part be related to the quality of responsiveness of attachment figures. Indeed, "good-enough" caregivers tend to mirror the infant's emotional expressions in a way that is both contingent (i.e., attuned to the infant's actual experience) and marked (i.e., signaling parent–infant distinctiveness) (Fonagy et al., 2002). Facial expressions have been shown to play an important role in this mirroring process (Beebe et al., 2016; Murray et al., 2016). The repeated experience of one's facial expressions being reliably met with a contingent and marked response from a caregiver is thought to foster the development of the infant's capacity for mentalizing and self-awareness, including awareness of their own emotional expressions and of how these expressions are perceived by others (Fotopoulou & Tsakiris, 2017; Rochat, 2009).

Individuals with BPD, however, often lacked such an emotionally responsive environment in early life (Agrawal et al., 2004; Crowell et al., 2009; Linehan, 1993). Several

studies suggest that their caregivers may have been inconsistent, responding at times in an intrusive and at times in a neglectful way to their emotional expressions (Lyons-Ruth et al., 1999; Lyons-Ruth et al., 2013). Both neglect and intrusiveness give rise to a mismatch between third-person objectification of the infant's experience and first-person experiences (Besharati et al., 2016), which may hamper the capacity to reflect on the (bodily) self from the perspective of the other (Neustadter, Fotopoulou, et al., 2019). Besides the caregiving environment, biological factors, such as high levels of impulsivity and/or poor effortful control, and their interaction, may further complicate this process (Gunderson & Lyons-Ruth, 2008; Luyten, Campbell, & Fonagy, 2020).

The Present Study

Several tasks have been developed to investigate transparency estimation (Gilovich et al., 1998). These tasks all have in common that they compare participants' estimations of the transparency of their own mental state (i.e., estimated transparency; ET) with the extent to which that mental state is in fact observable to others (i.e., objective transparency; OT). Studies in the general population have typically found a tendency to overestimate the transparency of one's own mental states, which has been called the "illusion of transparency" (Vorauer et al., 1999).

Given the impairments in third-person perspective taking and mentalizing described above, individuals with BPD may be even more prone to making errors in transparency estimation, in terms of both over- and underestimation, and their overestimation errors may be even larger than those seen in the general population. Yet, despite the obvious relevance of this body of research for our understanding of BPD, no studies to date have directly investigated transparency estimation in BPD.

This is the first study to investigate the capacity for transparency estimation in individuals high and low in BPD features. To this end, we adapted the experimental procedure of Barr and Kleck (1995) and showed participants emotion-eliciting movie clips while filming their reactive facial expressions. After each clip, participants rated how transparent they estimated their emotional experience to be while watching the particular clip (i.e., ET), as well as the intensity of their emotional experience (i.e., subjective arousal, SAR). The OT of their emotional experience was operationalized as the actual intensity of their facial emotional expressions, measured using automated computer vision facial coding software (Noldus FaceReader). Facial expressions are a rich source of social communication and are an important way by which inner mental states become outwardly visible to others, determining one's OT (Frith, 2009; Jack & Schyns, 2015). Aberrancies in the production of facial emotional expression have been found in individuals with BPD, but the findings are mixed, with some studies showing overall reduced facial responding (Herpertz et al., 2001; Renneberg et al., 2005) and others showing diminished positive and enhanced negative facial emotional expressions in these individuals (Matzke et al., 2014). However, no study to date has investigated awareness of the observability of facial emotional expressions in relation to BPD.

The following hypotheses were examined. First, we expected that individuals high in BPD features would be less accurate in estimating their own OT than those low in BPD features. We used two types of analyses to test this main research hypothesis. Based on discrepancy analyses, we expected that individuals high in BPD features would show less agreement between ET and OT than individuals low in BPD features. Second, we expected that ET would predict OT in individuals low in BPD features, but not, or to a lesser extent, in individuals high in BPD features. We tested this hypothesis using multilevel linear models while disaggregating two sources of variability in ET scores: intraindividual (i.e., variability

within the individual in response to the different film stimuli) and interindividual (i.e., variability between participants pooled over the observations for each participant). We expected to find that BPD group status moderates the association between ET and OT at both the within-subject level (i.e., impaired ability to correctly estimate within-person variability in transparency across different movie clips in individuals high in BPD features) and the between-subject level (i.e., impaired ability to estimate one's overall level of transparency in individuals high in BPD features).

Second, we aimed to explore the possible moderating role of SAR in the relationship between ET and OT. Moderate levels of SAR may relate to a better capacity to estimate transparency, whereas high levels of arousal (reflecting high self-salience of the emotional state) may lead individuals with BPD features to overestimate the transparency of emotional states, whereas low levels of arousal may lead these individuals to underestimate the transparency of those emotional states. Indeed, owing to a reduced capacity of individuals with BPD features to reflect on the (bodily) self from the perspective of others, the self-salience of the experience may determine whether these individuals will believe that this experience is shared by others. We expected to find a quadratic relationship between SAR and OT and ET in the high-BPD group, at both the within-subject and between-subject levels of transparency estimation, although, given the small sample size, these analyses should be considered exploratory.

Third, we hypothesized that SAR would show less correspondence with actual facial behavior in individuals with BPD features compared with those without BPD, as individuals with BPD may have difficulty identifying and labeling their own emotional states (Derks et al., 2017). We thus expected that SAR would predict OT in the low-BPD group but not in the high-BPD group.

Fourth, as argued above, we expected higher levels of SAR during the task to correlate with overestimating transparency, and lower levels of SAR to correlate with underestimating transparency. Fifth, we expected impairments in mentalizing (as measured using the Reflective Functioning Questionnaire (RFQ)) to correlate with transparency estimation in two distinct ways: whereas extreme certainty about the mental states of self and others may lead one to overestimate transparency, extreme uncertainty about mental states was expected to be correlated with underestimating transparency. Finally, childhood trauma and attachment insecurity were expected to be positively correlated with both types of transparency errors, as they may reflect a history of inconsistent (both neglectful and intrusive) parental responsiveness to one's facial emotional expressions.

Methods

Participants and Procedure

The Maclean Screening Instrument for Borderline Personality Disorder (MSI-BPD) (Zanarini et al., 2003) was administered to a large sample (N = 397; 90% female) of first-year psychology students at a large university in Belgium in exchange for course credits. Female students scoring above the clinical cut-off for BPD (scores >6 out of 10) (high-BPD group) and individuals with low scores (0 or 1 out of 10) (low-BPD group) were invited to participate in a larger study investigating self–other distinction in relation to BPD, which included the present experiment. A total of n = 62 individuals (age range = 17–20 years) participated in the study, of whom 32 were in the low-BPD group and 30 were in the high-BPD group. The sample size was determined by an a priori power analysis (n = 58 needed to detect a between-factor effect (medium size, f = 0.25) in an ANOVA with 5 repeated measures and 2 groups with a power of 0.8, calculated using Gpower 3.1 (Faul et al., 2007)). The observed power of the present analyses was, however, potentially lower than 0.8, as the correlations between

repeated measures were higher than the assumed 0.3 (r = 0.56–0.70 between repeated measures of OT; r = 0.36–0.60 between repeated measures of ET), and due to testing of complex/parameter-heavy interactions.

The study involved a 1-hour session in a research room at the university. The participants first completed a set of questionnaires (described below) and then took part in the transparency estimation task, which lasted about 25 minutes. Both the questionnaires and the experimental tasks were presented on a laptop using Qualtrics (Provo, UT, USA). The participants' faces were filmed using a 10-megapixel camera (Panasonic HC-V180) using a standard set-up so that distance and lighting were standardized. Participants were informed that they would be filmed during the experiment. The study was approved by the Social and Societal Ethics Committee of the University of Leuven. Written informed consent was obtained before the start of the experiment.

Measures

Maclean Screening Inventory for BPD (MSI-BPD). The MSI-BPD (Zanarini et al., 2003) comprises 10 yes-or-no items based on DSM-IV BPD criteria. A score of 7 or higher was found to be the best cut-off in providing both good sensitivity (0.81) and specificity (0.85) in separating individuals with and without BPD (Zanarini et al., 2003). There is evidence for the concurrent and convergent validity of the MSI-BPD in a nonclinical population (Gardner & Qualter, 2009) and in a predominantly female undergraduate population (Verschuere & Tibboel, 2011). The internal consistency of the MSI-BPD was excellent in the current sample (α = 0.92).

Experiences in Close Relationships – Revised (ECR-R). The ECR-R has two subscales, comprising 18 items each, that measure attachment-related anxiety and avoidance (Fraley et al., 2000). The two-factor structure of the ECR-R was replicated (Kooiman et al.,

2013; Sibley et al., 2005). The reliability of the anxiety (α = 0.90) and the avoidance (α = 0.94) subscales was excellent in the present study.

Reflective Functioning Questionnaire (RFQ). The RFQ is a self-report measure of mentalizing that includes eight items answered on a Likert scale that ranges from 1 (*I do not agree at all*) to 6 (*I very much agree*). The items from the Uncertainty subscale of the RFQ (RFQ-U) are rescored so that high scores reflect extreme uncertainty about mental states (e.g., "*I don't always know why I do what I do*"), assumed to reflect hypomentalizing. The items from the Certainty subscale (RFQ-C) are also rescored, so that higher scores reflect more certainty about mental states (e.g., "*I always know what I feel"*), with very high scores reflecting excessive mentalizing or hypermentalizing. The RFQ-U scale has been found to relate to other measures assessing impairments in mentalizing and to measures of personality pathology, while the RFQ-C has been shown to relate to conditions associated with hypermentalizing, for example, in patients with eating disorder features (Badoud et al., 2015; Fonagy et al., 2016). Cronbach's alpha in the current study was good for the RFQ-C (α = 0.73) but poor for the RFQ-U (α = 0.53).

Childhood Trauma Questionnaire – Short Form (CTQ-SF). The CTQ-SF (Bernstein et al., 2003) assesses the experience of childhood maltreatment with five subscales of five items each: physical abuse (PA), emotional abuse (EA), sexual abuse (SA), physical neglect (PN), and emotional neglect (EN). In this study, we used the CTQ Total scale, which had excellent internal consistency (α = 0.89).

Transparency Estimation task. We adapted the experimental design by Barr and Kleck (1995) to measure transparency estimation. Participants watched five two-minute film clips, which were selected scenes from feature films with attachment-related themes, that may elicit a strong emotional experience (Hewig et al., 2005). The clips were shown in a fixed order, with the four negative clips shown first and the positive clip last. Each film clip was

preceded by a short description of the context of the clip to ensure a good understanding of the scene (see Supplementary Table S1). After each film clip, participants rated their emotional arousal 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 4 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*"). The items were piloted to ensure they were understood correctly.

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). The software analyzes the facial expression at a rate of 15 frames per second using an artificial neural network trained with over 10,000 manually annotated images, yielding intensity scores per frame between 0 and 1, with 0 representing 0% agreement between the frame and the database images and 1 representing a 100% match, for each emotion (happy, sad, angry, surprised, scared, disgusted, and contemptuous) as well as a neutral expression score. 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 scores on the seven different emotion categories were summed per participant per clip to create 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 accuracy for emotion classification between 80% and 88% (Lewinski et al., 2014; Skiendziel et al., 2019).

Data Analyses

Multilevel linear models (MLMs) were used to analyze the data with the *lme4* package (Bates et al., 2015) in R (version 3.6). First, we analyzed group differences and the impact of the different film clips on ET, OT, and SAR (see Supplementary Table S2 and Figure S1). Differences between groups and between film clips were further examined with Tukey-adjusted post-hoc tests.

Second, we analyzed differences between groups and between film clips in the discrepancy between standardized ET and standardized OT scores, as well as in the absolute value of this discrepancy. The data were then split between observations with overestimation (discrepancy score >0) and observations with underestimation (discrepancy score <0) to test for group differences.

Third, we ran a series of MLMs predicting OT based on both the within- and the between-subject level variability and interactions of ET, SAR, and BPD group status. A random intercept per participant was added to account for the fact that five repeated measures were nested within participants. The time-varying predictors ET and SAR were disaggregated in a within-person centered (wpc) and a between-person centered (bpc) part to assess both sources of information separately instead of confounding them (Curran & Bauer, 2011). The within-person centered part of the predictors was created by centering each score around each participant's person-mean, capturing the amount by which a participant deviated from her own average at each film clip. The between-person centered part, calculated by centering each person-mean around the grand mean (i.e., the mean of all person-means in the total sample), reflects on average how much a person differs from others in the sample on that predictor variable (Howard, 2015). No detrending (i.e., removal of time effects) of the time-varying variables OT, ET, and SAR was applied because the time effect was caused by the

experimental manipulation of emotion elicitation using different film stimuli (Wang & Maxwell, 2015).

A MLM was built including main effects and two-way interactions between Group and ET.bpc and Group and ET.wpc, resulting in Model 1 presented in Supplementary Table S3. Next, we added main effects and interactions with the linear terms SAR.wpc and SAR.bpc to test whether SAR moderated the association between ET and OT. From this model, the three-way interactions Group × ET.wpc × SAR.wpc ($\chi^2(2) = 1.9$, $p = 0.39$) and Group × ET.bpc × SAR.bpc ($\chi^2(2) = 0.11$, $p = 0.94$) were excluded because they did not improve the model fit, resulting in the final Model 2 reported in Supplementary Table S3. Model 2 provided a significantly better fit to the data than Model 1 ($\chi^2(8) = 33.36$, $p < 0.001$). Finally, the main effects and interactions with the quadratic terms SAR.wpc² and SAR.bpc² were included to test the hypothesized quadratic effect of SAR on the association between ET and OT (see Table S3). However, none of the quadratic terms were significant and Model 3 did not provide a better fit to the data than Model 2 ($\chi^2(12) = 20.31$, $p = 0.06$). Model 2 was thus retained as the final model and is presented in Table 2.

The data from one participant low in BPD features could not be used because the FaceReader calibration procedure failed, and six more observations from three participants were missing due to technical issues with their face recordings, yielding a total of 299 observations from 61 unique participants (3.54% missing data). The maximum likelihood estimation method used in the MLMs is well suited to handle small amounts of missing data (Molenberghs & Kenward, 2007). Outliers were present in the MLMs. Neither the removal of these outliers nor the application of a robust MLM method (Koller, 2016) yielded substantive differences compared with the original analyses (see Supplementary Table S4). Therefore, all analyses were reported on the full dataset.

Group differences in questionnaire data were investigated with independent sample *t*-tests. Associations between questionnaire measures and the under- and overestimation of transparency (calculated per film clip as the standardized ET score minus the standardized OT score and then averaged across clips) as well as the general error rate in transparency estimation (i.e., the averaged value of the absolute difference scores) were analyzed using Pearson correlations. These analyses were performed with IBM SPSS statistics for Windows, version 25 (IBM Corp., Armonk, NY, USA).

Results

Sample Description

The high-BPD and low-BPD groups did not differ in age (see Table 1). Psychotropic medication use was low, with only two participants (3.2% of total sample) of the high-BPD group reporting the use of either a sleep or an antidepressant medication. The groups differed significantly from each other on all questionnaire measures; specifically, the high-BPD group scored higher than the low-BPD group on attachment avoidance and anxiety, childhood trauma, and uncertainty about mental states. In contrast, the low-BPD group showed higher levels of certainty about mental states.

[TABLE 1 ABOUT HERE]

Transparency Estimation

Overall, there was a significant main effect of the different film clips on ET, OT, and SAR, but there were no significant main effects of BPD group status (see Supplementary Table S2 and Figure S1). Specifically, Tukey-adjusted post-hoc tests showed that ET scores were lower in response to the first clip, and higher in response to the positive clip that was shown last. Furthermore, OT and SAR were higher in the positive clip than in the negative

clips, although the high-BPD group reported significantly lower SAR in response to the positive clip compared with the low-BPD group.

The discrepancy between standardized ET and standardized OT scores per clip was calculated, with higher positive scores indicating transparency underestimation (ET > OT), and more negative scores indicating transparency underestimation (ET < OT). MLMs showed no significant group difference on this measure (B = –0.20, 95% CI [–0.77; 0.37], $p = 0.5$). However, in line with our hypothesis, the absolute value of the discrepancy score was significantly higher in the high-BPD group than the low-BPD group (B = 0.38, 95% CI [0.11; 0.65], $p = 0.008$), showing that individuals in the high-BPD group made more errors when estimating their transparency, expressed in both directions. When looking at over- and underestimation separately, the high-BPD group was found to show significantly larger underestimation errors than the low-BPD group (B = –0.44, 95% CI [–0.78; –0.11], $p = 0.012$) and larger overestimation errors, but this difference was not significant (B = 0.24, 95% CI [–0.10; 0.58], $p = 0.17$). There were no main effects or interactions with "Film".

[TABLE 2 ABOUT HERE]

The results from the linear analysis presented in Table 2 show that, in line with the first hypothesis, between-person variability in ET predicted OT in the low-BPD group, but not in the high-BPD group, as shown by the significant Group × ET.bpc interaction. As Figure 1 shows, whereas interindividual differences in ET were related to OT in the low-BPD group (e.g., individuals experiencing a high level of transparency also effectively showed higher intensity levels of facial expressions), this association was not present in the high-BPD group. Contrary to our expectations, within-person variability in ET predicted OT in both groups, as the Group × ET.wpc interaction was not significant.

[FIGURE 1 ABOUT HERE]

Contrary to our second hypothesis, moderate levels of SAR did not predict better transparency estimation compared with high and low levels of SAR, as there was no quadratic effect of SAR on the association between ET and OT (see Supplementary Table S3). Instead, as shown in the model presented in Table 2, the association between within-person variability in ET and OT was moderated by BPD group status and SAR linearly, as shown by the significant Group × ET.wpc × SAR.bpc interaction. Specifically, individuals in the low-BPD group were objectively more transparent in those clips where they estimated being more transparent than in other clips, regardless of SAR (see Figure 2, left panel). Individuals in the high-BPD group reporting on average high levels of arousal relative to others (higher SAR.bpc) were, however, superior at estimating variability in their own transparency across the different clips, and those reporting low arousal were inferior, compared with the low-BPD group (see Figure 2, right panel).

[FIGURE 2 ABOUT HERE]

Moreover, in both groups, the quality of between-subject level transparency estimation was moderated by intra-individual variability in SAR, as the ET.bpc × SAR.wpc interaction was significant. This result shows that interindividual differences in ET level better predicted OT in those film clips where participants reported feeling more intense emotion than in other clips, although between-subject level transparency estimation remained impaired in the high-BPD group regardless of SAR.

Finally, the significant Group × SAR.wpc interaction confirmed the third hypothesis. Indeed, individuals in the low-BPD group were objectively more transparent in response to those clips where they reported higher arousal than in other clips, whereas in the high-BPD group, within-person variability in SAR did not predict OT.

[FIGURE 3 ABOUT HERE]

Associations with Attachment, Mentalizing, and Childhood Trauma

OT, ET, and SAR during the task were not significantly associated with questionnaire measures in the total sample. However, in the high-BPD group only, SAR was associated with certainty about mental states (RFQ-C) ($r = 0.42$, $p = 0.02$), and in the low-BPD group, attachment avoidance was related to lower ET ($r = –0.38$, $p = 0.03$). As expected, in the total sample, overestimating transparency related significantly to higher RFQ-C, and on a trend-level to more SAR (see Table 3). In the low-BPD group only, attachment avoidance and childhood trauma were related to underestimating transparency. Other measures did not relate to over- or underestimation of transparency specifically, but instead, they related to a general tendency to make errors in transparency estimation in both directions. Indeed, a higher incidence of BPD features, attachment anxiety, and uncertainty about mental states (RFQ-U) were significantly associated with more errors on transparency estimation (both over- and underestimation).

[TABLE 3 ABOUT HERE]

Discussion

The present study investigated, for the first time, the capacity for transparency estimation in nonclinical participants high and low in BPD features. Participants estimated the transparency of their emotional experience while watching a series of emotion-eliciting movie clips, and the OT of their emotional experience was determined using facial expression coding software (Noldus FaceReader). Individuals high in BPD features showed a larger discrepancy between ET and OT compared with participants low in BPD features, in terms of both overestimating (ET > OT) and underestimating (ET < OT) transparency. These findings show that, whereas individuals without BPD features tended to slightly overestimate transparency, in line with previous findings in the general population (Vorauer et al., 1999;

Vorauer & Cameron, 2002), some individuals high in BPD features overestimated their transparency to an even larger degree than individuals without BPD features, while others significantly underestimated the transparency of their emotions.

Moreover, individuals without BPD features who experienced high levels of transparency across all clips effectively showed higher intensity levels of facial expressions, whereas this association did not exist in the group of participants high in BPD features. In other words, individuals in the high-BPD group who estimated their transparency to be high were not objectively more transparent, indicative of transparency overestimation. Similarly, those who judged the transparency of their emotional experience to be low did not exhibit less intense facial emotion expressions during the task, showing transparency underestimation. This shows that individuals high in BPD features are less able to gauge how much of their emotional experience is observable to others through their facial expressions, or, in other words, to imagine their own facial expressions from another person's perspective. This is in line with previous studies showing impairments in perspective taking in individuals with BPD (Colle et al., 2018; New et al., 2012); however, this is the first study showing impairments in perspective taking with regard to the (embodied) self in relation to BPD.

Contrary to what we hypothesized, moderate levels of SAR did not predict a better capacity for transparency estimation compared with high and low levels of arousal, although this may be due to fact that the statistical power of our study was insufficient to test this complex quadratic interaction. Instead, SAR moderated transparency estimation linearly. Indeed, on the within-subject level of transparency estimation, the ability to correctly estimate intraindividual variability in transparency across the different film clips was disturbed only in those participants high in BPD features who reported overall low arousal during the task. Those reporting high arousal were even superior to individuals without BPD features in estimating how detectable changes in their own facial expressiveness would be to others. An

explanation for these results may be found in the relationship between interoception and self-attention in BPD. The ability to estimate changes in one's own facial behavior may, in part, rely on the processing of afferent information from the facial muscles and thus on interoception (Löffler et al., 2018). Interestingly, whereas deficits in the neural representation of interoceptive signals have been reported in individuals with BPD (Müller et al., 2015; Schmitz et al., 2020), another study reported that BPD patients' interoceptive accuracy was intact when they were instructed to pay attention to their bodily signals (Hart et al., 2012). Against this background, higher self-reported arousal during the task may reflect increased self-salience or self-attention, which may result in enhanced interoceptive accuracy in the high-BPD group, whereas those reporting low arousal may default to impaired interoceptive performance. Furthermore, the fact that increased self-reported arousal related to more extreme certainty about mental states (RFQ-C) in the high-BPD group suggests that excessive mentalizing (or hypermentalizing) may relate to hyper-awareness of one's own bodily signals in BPD. This is consistent with clinical observations (Sharp & Vanwoerden, 2015), but further research is needed.

Between-subject level transparency estimation (i.e., the ability to estimate one's overall level of transparency relative to others) was disturbed in the high-BPD group regardless of the level of self-reported arousal. This shows that although increased self-salience in the high-BPD group may increase awareness of changes in their own facial expressions, they may not be able to shift from this self-focus to imagine the self from a third-person perspective. This points to an impairment in flexibly shifting between representing "self" and representing "others" in those with BPD (Lamm et al., 2016b; Quesque & Brass, 2019; Sowden & Shah, 2014), and provides further evidence for a self–other distinction impairment in BPD. Self–other distinction impairment has been found on an embodied level, with BPD patients having difficulty distinguishing their own and others' actions (Hauschild et

al., 2018) and bodily representations (Bekrater-Bodmann et al., 2016; De Meulemeester et al., 2020; Neustadter, Fineberg, et al., 2019). On the level of mental states, previous research has shown that individuals with BPD have difficulty mentalizing the mind of others as separate from their own (Colle et al., 2018), and with representing their own and multiple others' perspectives (Petersen et al., 2016). Our findings show that the self–other distinction impairment may also be expressed in terms of difficulties mentalizing the "self" from the perspective of others (so-called personal second-order mentalizing (Wu et al., 2020)).

Moreover, intraindividual variability in SAR across the task predicted OT in the low-BPD group, but not in the high-BPD group. In other words, the high-BPD group did not express more emotion in their faces during those clips where they reported feeling more intense emotion and vice versa. On the one hand, this dissociation between emotional experience and facial emotion expressions may point toward impaired facial feedback on emotional experience. Indeed, proprioceptive awareness of one's own facial expressions helps to identify one's emotional states (Wood et al., 2016), however, this mechanism may be less effective in individuals with BPD, resulting in low emotional awareness or alexithymia (Derks et al., 2017). On the other hand, it may point toward aberrant facial emotion expression in those with BPD (Renneberg et al., 2005; Staebler et al., 2011), in the sense that their facial expressions are less aligned with their inner experience compared with individuals without BPD. This decoupling of emotional experience and facial emotion expression may make it difficult for others to mentalize about the inner mental states of individuals with BPD based on their facial expressions. Indeed, Flury et al. (2008) found that BPD patients were more difficult to "read" by others than controls.

Interestingly, during this task, the high-BPD and low-BPD groups did not differ in terms of mean levels of self-reported emotional arousal, and OT and ET. However, results clearly indicate that it is the association between the experience of emotion, the expression of

the emotion, and the higher-order representation of the emotion that was disturbed in individuals high in BPD features, pointing toward an impairment in the cross-modal integration of different mentalizing abilities (Luyten, Campbell, Allison, et al., 2020).

Impaired transparency estimation in individuals high in BPD features was expressed in two directions, as a tendency to both overestimate (ET > OT) and underestimate (ET < OT) transparency. Attachment anxiety was the strongest correlate of the tendency to both over- and underestimate transparency, which was driven by the high-BPD group. Attachment anxiety relates to disturbed mirroring of facial affect in parent–child interactions (Kim et al., 2014; Shai & Belsky, 2017). Thus, anxiously attached individuals may have not received reliable feedback on their facial affect displays by attachment figures, which may result in the impaired capacity to imagine their own facial behavior through the eyes of others. For the same reasons, we expected childhood trauma to be related to making transparency errors, but this correlation was apparent only at trend level, possibly due to the fact that the prevalence of childhood trauma was relatively low in this sample (e.g., compared with the clinical and nonclinical sample means in Thombs et al. (2009)). Contrary to our hypotheses, attachment avoidance was unrelated to transparency estimation in the high-BPD group and related to transparency underestimation in the low-BPD group. Transparency estimation was significantly related to mentalizing, although the correlations were weak and should be interpreted with caution given the low internal consistency of the RFQ, in particular the RFQ-U subscale. Certainty about mental states (RFQ-C), which in its extremes may reflect excessive mentalizing or hypermentalizing, related to transparency overestimation. Indeed, individuals who hypermentalize may be so certain that others will know what they know and feel that they fail to recognize that mental states are largely opaque unless they are communicated (Bateman & Fonagy, 2013; Bo et al., 2017). Extreme uncertainty about mental

states of self and others (RFQ-U), on the other hand, related to making transparency errors in both directions.

Clinicians have noted that patients with BPD often misjudge how others interpret their facial expressions (Dhaliwal et al., 2020). Our findings show that these misjudgments may be expressed in terms of both over- and underestimating the intensity of their facial emotion expressions. They may, for instance, assume that others will be able to know how angry they feel (while keeping a still face), which may lead to misunderstandings (Bo et al., 2017; Cameron & Vorauer, 2008), or they may overestimate the extent to which others can see how anxious they are, which may exacerbate anxiety (Savitsky & Gilovich, 2003) or lead to increased self-conscious emotions such as shame (Thomas et al., 2018). On the other hand, they may be unaware of how much emotion is expressed in their faces (transparency underestimation), which may hamper their ability to gauge how they are perceived by others. In addition, the emotion they express in their face without recognizing it may be automatically mimicked by interaction partners (Dimberg et al., 2000; Olszanowski et al., 2019), who may become confused about the source of the emotion and wrongly attribute the emotional state to themselves. In this way, transparency underestimation may lead to projective identification, a defense mechanism often associated with BPD (Zanarini et al., 2013).

It can be argued that current evidence-based treatments for BPD often involve a focus on making transparency assumptions explicit so that the patient can become aware of the mismatch between their perceived and actual nonverbal emotion communication. Indeed, face-to-face patient–therapist contact may target BPD patients' errors in transparency estimation at two levels: (1) implicitly, by providing contingent and marked mirroring of the patient's facial expressions, and (2) explicitly, through mentalizing together about the patient's transparency assumptions and discussing how the therapist actually perceives the patient's facial expressions. Mentalization-based treatment (Bateman & Fonagy, 2013) and

transference-focused psychotherapy (Levy et al., 2019) combine this implicit and explicit focus through mentalizing together about the patient's nonverbal communications and through emphasizing the process of marked mirroring in the treatment. Dialectical behavior therapy focuses on fostering patients' understanding of emotions at an embodied level and on how these are communicated to others (May et al., 2016). It remains, however, an empirical question whether such interventions are related to improved transparency estimation and to treatment outcome.

Several limitations need to be considered when interpreting the findings of this study. The first limitation is the inclusion of a nonclinical sample of students high and low in BPD features as characterized by a self-report screen. Replication in carefully diagnosed clinical BPD patients and matched controls is necessary to validate these findings. Yet, there is increasing evidence that supports a more dimensional rather than categorical approach to BPD (Kotov et al., 2017), and even subthreshold BPD symptoms have been found to impact psychological functioning in young people (Chanen et al., 2017). Second, only female students were included to avoid potential confounding effects of gender; however, this precludes the generalization of the results to men with BPD. The third important limitation of the present study is its small sample size. Although the sample size was based on a priori power analysis, to detect significant interactions, especially with regard to the more complex/parameter-heavy interactions, such as the quadratic analyses, a large sample size is required. Hence, further studies in larger samples are needed. Fourth, a part of the experiment relied on self-report. Finally, although facial expressions determine a large part of one's OT, participants may have based their transparency estimation on other nonverbal signals that were not taken into account, such as posture, head pose, or blushing.

In conclusion, the present study provides the first demonstration of impaired transparency estimation in individuals with BPD features, showing that individuals high in

BPD features are less able to represent their (embodied) selves from the perspective of others. These findings provide further evidence for problems in integrating information based on different mentalizing capacities in BPD, and in self–other distinction in particular.

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

American Psychiatric Association. (2013). Diagnositic and statistical manual of mental disorders (5th ed.). *Arlington, VA: American Psychiatric Publishing.* https://doi.org/10.1176/appi.books.9780890425596.744053

Badoud, D., Luyten, P., Fonseca-Pedrero, E., Eliez, S., Fonagy, P., & Debbané, M. (2015). The French Version of the Reflective Functioning Questionnaire: Validity Data for Adolescents and Adults and Its Association with Non-Suicidal Self-Injury. *PLoS One, 10*, e0145892. https://doi.org/10.1371/journal.pone.0145892

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

Bateman, A., & Fonagy, P. (2013). Mentalization-Based Treatment. *Psychoanalytic Inquiry, 33*(6), 595-613. https://doi.org/10.1080/07351690.2013.835170

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

Beeney, Hallquist, M. N., Ellison, W. D., & Levy, K. (2016). Self-other disturbance in borderline personality disorder: Neural, self-report, and performance-based evidence. *Journal of Personality Disorders*, *7*, 28-39. https://doi.org/10.1037/per0000127

Bekrater-Bodmann, R., Chung, B. Y., Foell, J., Gescher, D. M., Bohus, M., & Flor, H. (2016). Body plasticity in borderline personality disorder: A link to dissociation. *Comprehensive Psychiatry*, *69*, 36-44. https://doi.org/10.1016/j.comppsych.2016.05.002

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

Besharati, S., Forkel, S. J., Kopelman, M., Solms, M., Jenkinson, P. M., & Fotopoulou, A. (2016). Mentalizing the body: spatial and social cognition in anosognosia for hemiplegia. *Brain*, *139*(Pt 3), 971-985. https://doi.org/10.1093/brain/awv390

Bo, S., Sharp, C., Fonagy, P., & Kongerslev, M. (2017). Hypermentalizing, Attachment, and Epistemic Trust in Adolescent BPD: Clinical Illustrations. *Personality Disorders-Theory Research and Treatment*, *8*(2), 172-182. https://doi.org/10.1037/per0000161

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

Chambers, J. R., Epley, N., Savitsky, K., & Windschitl, P. D. (2008). Knowing too much: using private knowledge to predict how one is viewed by others. *Psychological Science, 19*(6), 542-548. https://doi.org/10.1111/j.1467-9280.2008.02121.x

Chanen, A., Sharp, C., & Hoffman, P. (2017). Prevention and early intervention for borderline personality disorder: a novel public health priority. *World psychiatry : official journal of the World Psychiatric Association (WPA), 16*(2), 215-216. https://doi.org/10.1002/wps.20429

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., Panagiotopoulou, E., Fotopoulou, A., & Luyten, P. (2020). Self–other distinction and borderline personality disorder features: Evidence for egocentric and altercentric bias in a self–other facial morphing task. *Personality Disorders-Theory Research and Treatment*, Advance online publication. https://doi.org/10.1037/per0000415

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

Dhaliwal, K., Danzig, A., & Fineberg, S. K. (2020). Improving Research Practice for Studying Borderline Personality Disorder: Lessons From the Clinic. *Chronic Stress, 4*, 2470547020912504. https://doi.org/10.1177/2470547020912504

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

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

Flury, J. M., Ickes, W., & Schweinle, W. (2008). The borderline empathy effect: Do high BPD individuals have greater empathic ability? Or are they just more difficult to "read"? *Journal of Research in Personality, 42*(2), 312-332. https://doi.org/10.1016/j.jrp.2007.05.008

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

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1535), 3453-3458. https://doi.org/10.1098/rstb.2009.0142

Gardner, K., & Qualter, P. (2009). Reliability and validity of three screening measures of borderline personality disorder in a nonclinical population. *Personality and Individual Differences, 46*(5-6), 636-641. https://doi.org/10.1016/j.paid.2009.01.005

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

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

Gunderson, J. G., & Lyons-Ruth, K. (2008). BPD's interpersonal hypersensitivity phenotype: a gene-environment-developmental model. *Journal of Personality Disorders, 22*(1), 22-41. https://doi.org/10.1521/pedi.2008.22.1.22

Harari, H., Shamay-Tsoory, S. G., Ravid, M., & Levkovitz, Y. (2010). Double dissociation between cognitive and affective empathy in borderline personality disorder. *Psychiatry Research, 175*(3), 277-279. https://doi.org/10.1016/j.psychres.2009.03.002

Hart, N., McGowan, J., Minati, L., & Critchley, H. D. (2012). Emotional Regulation and Bodily Sensation: Interoceptive Awareness Is Intact in Borderline Personality Disorder. *Journal of Personality Disorders*, *27*, 1-13. https://doi.org/10.1521/pedi_2012_26_049

Hauschild, S., Winter, D., Thome, J., Liebke, L., Schmahl, C., Bohus, M., & Lis, S. (2018). Behavioural mimicry and loneliness in borderline personality disorder. *Comprehensive Psychiatry*, *82*, 30-36. https://doi.org/10.1016/j.comppsych.2018.01.005

Herpertz, S. C., Bertsch, K., & Jeung, H. (2018). Neurobiology of Criterion A: self and interpersonal personality functioning. *Current Opinion in Psychology*, *21*, 23-27. https://doi.org/10.1016/j.copsyc.2017.08.032

Herpertz, S. C., Jeung, H., Mancke, F., & Bertsch, K. (2014). Social dysfunctioning and brain in borderline personality disorder. *Psychopathology*, *47*, 417-424. https://doi.org/10.1159/000365106

Herpertz, S. C., Werth, U., Lukas, G., Qunaibi, M., Schuerkens, A., Kunert, H. J., Freese, R., Flesch, M., Mueller-Isberner, R., Osterheider, M., & Sass, H. (2001). Emotion in criminal offenders with psychopathy and borderline personality disorder. *Archives of General Psychiatry*, *58*(8), 737-745. https://doi.org/DOI 10.1001/archpsyc.58.8.737

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

Howard, A. L. (2015). Leveraging Time-Varying Covariates to Test Within- and Between-Person Effects and Interactions in the Multilevel Linear Model. *Emerging Adulthood*, *3*(6), 400-412. https://doi.org/10.1177/2167696815592726

Jack, Rachael E., & Schyns, Philippe G. (2015). The Human Face as a Dynamic Tool for Social Communication. *Current Biology, 25*(14), R621-R634. https://doi.org/https://doi.org/10.1016/j.cub.2015.05.052

Kim, S., Fonagy, P., Allen, J., Martinez, S., Iyengar, U., & Strathearn, L. (2014). Mothers who are securely attached in pregnancy show more attuned infant mirroring 7 months postpartum. *Infant Behavior and Development, 37*(4), 491-504. https://doi.org/https://doi.org/10.1016/j.infbeh.2014.06.002

Koller, M. (2016). robustlmm: An R Package for Robust Estimation of Linear Mixed-Effects Models. *2016, 75*(6), 24. https://doi.org/10.18637/jss.v075.i06

Kooiman, C. G., Klaassens, E. R., Lugt, J. Q. V., & Kamperman, A. M. (2013). Psychometrics and Validity of the Dutch Experiences in Close RelationshipsRevised (ECRr) in an Outpatient Mental Health Sample. *Journal of Personality Assessment, 95*(2), 217-224. https://doi.org/10.1080/00223891.2012.740540

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., Miller, J. D., Moffitt, T. E., Morey, L. C., Mullins-Sweatt, S. N., Ormel, J., Patrick, C. J., Regier, D. A., Rescorla, L., Ruggero, C. J., Samuel, D. B., Sellbom, M., Simms, L. J., Skodol, A. E., Slade, T., South, S. C., Tackett, J. L., Waldman, I. D., Waszczuk, M. A., Widiger, T. A., Wright, A. G. C., & Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology, 126*(4), 454-477. https://doi.org/10.1037/abn0000258

Lamm, C., Bukowski, H., & Silani, G. (2016a). From shared to distinct self-other representations in empathy: evidence from neurotypical function and socio-cognitive

disorders. *Philosophical Transactions of the Royal Society B-Biological Sciences, 371*(1686). https://doi.org/10.1098/Rstb.2015.0083

Lamm, C., Bukowski, H., & Silani, G. (2016b). From shared to distinct self–other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*, 20150083. https://doi.org/10.1098/rstb.2015.0083

Levy, K. N., Draijer, N., Kivity, Y., Yeomans, F. E., & Rosenstein, L. K. (2019). Transference-Focused Psychotherapy (TFP). *Current Treatment Options in Psychiatry, 6*(4), 312-324. https://doi.org/10.1007/s40501-019-00193-9

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

Linehan, M. (1993). *Cognitive-behavioral Treatment of Borderline Personality Disorder*. The Guildford Press.

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. *Annu Rev Clin Psychol*. https://doi.org/10.1146/annurev-clinpsy-071919-015355

Luyten, P., Campbell, C., & Fonagy, P. (2020). Borderline personality disorder, complex trauma, and problems with self and identity: A social-communicative approach. *Journal of Personality, 88*(1), 88-105. https://doi.org/10.1111/jopy.12483

Luyten, P., De Meulemeester, C., & Fonagy, P. (2020). The self–other distinction in psychopathology: Recent developments from a mentalizing perspective. In M. Gilead & K. N. Ochsner (Eds.), *The Neural Basis of Mentalizing*. Springer International Publishing.

Luyten, P., & Fonagy, P. (2015). The neurobiology of mentalizing. *Personality Disorders: Theory, Research, and Treatment, 6*, 66-379. https://doi.org/10.1037/per0000117

Lyons-Ruth, K., Bronfman, E., & Parsons, E. (1999). Atypical attachment in infancy and early childhood among children at developmental risk. IV. Maternal frightened, frightening, or atypical behavior and disorganized infant attachment patterns. *Monographs of the Society for Research in Child Development, 64*(3), 67. https://doi.org/10.1111/1540-5834.00034.

Lyons-Ruth, K., Bureau, J.-F., Holmes, B., Easterbrooks, A., & Brooks, N. H. (2013). Borderline symptoms and suicidality/self-injury in late adolescence: prospectively observed relationship correlates in infancy and childhood. *Psychiatry Research, 206*(2-3), 273-281. https://doi.org/10.1016/j.psychres.2012.09.030

Matzke, B., Herpertz, S. C., Berger, C., Fleischer, M., & Domes, G. (2014). Facial reactions during emotion recognition in borderline personality disorder: a facial electromyography study. *Psychopathology, 47*(2), 101-110. https://doi.org/10.1159/000351122

May, J. M., Richardi, T. M., & Barth, K. S. (2016). Dialectical behavior therapy as treatment for borderline personality disorder. *Mental Health Clinician, 6*(2), 62-67. https://doi.org/10.9740/mhc.2016.03.62

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

Müller, L. E., Schulz, A., Andermann, M., Gäbel, A., Gescher, D. M., Spohn, A., Herpertz, S. C., & Bertsch, K. (2015). Cortical Representation of Afferent Bodily Signals in Borderline Personality Disorder: Neural Correlates and Relationship to Emotional Dysregulation. *Jama Psychiatry, 72*, 1-10. https://doi.org/10.1001/jamapsychiatry.2015.1252

Murray, L., De Pascalis, L., Bozicevic, L., Hawkins, L., Sclafani, V., & Ferrari, P. F. (2016). The functional architecture of mother-infant communication, and the development of infant social expressiveness in the first two months. *Scientific Reports, 6*, 39019-39019. https://doi.org/10.1038/srep39019

Neustadter, E. S., Fineberg, S. K., Leavitt, J., Carr, M. M., & Corlett, P. R. (2019). Induced illusory body ownership in borderline personality disorder. *Neuroscience of Consciousness, 2019*(1), niz017. https://doi.org/10.1093/nc/niz017

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

Olszanowski, M., Wrobel, M., & Hess, U. (2019). Mimicking and sharing emotions: A re-examination of the link between facial mimicry and emotional contagion. *Cognition and Emotion, 34*. https://doi.org/10.1080/02699931.2019.1611543

Peters, J. R., & Geiger, P. J. (2016). Borderline Personality Disorder and Self-Conscious Affect: Too Much Shame but Not Enough Guilt? *Personality Disorders-Theory Research and Treatment*, *7*(3), 303-308. https://doi.org/10.1037/per0000176

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

Quesque, F., & Brass, M. (2019). The Role of the Temporoparietal Junction in Self-Other Distinction. *Brain Topography*, *32*(6), 943-955. https://doi.org/10.1007/s10548-019-00737-5

Renneberg, B., Heyn, K., Gebhard, R., & Bachmann, S. (2005). Facial expression of emotions in borderline personality disorder and depression. *Journal of Behavior Therapy and Experimental Psychiatry*, *36*(3), 183-196. https://doi.org/10.1016/j.jbtep.2005.05.002

Ripoll, L. H., Snyder, R., Steele, H., & Siever, L. J. (2013). The neurobiology of empathy in borderline personality disorder. *Current Psychiatry Reports*, *15*, 344. https://doi.org/10.1007/s11920-012-0344-1

Rochat, P. (2009). *Others in mind: Social origins of self-consciousness*. Cambridge University Press. https://doi.org/10.1017/cbo9780511812484

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

Schmitz, M., Müller, L. E., Schulz, A., Kleindienst, N., Herpertz, S. C., & Bertsch, K. (2020). Heart and brain: Cortical representation of cardiac signals is disturbed in borderline personality disorder, but unaffected by oxytocin administration. *Journal of Affective Disorders*, *264*, 24-28. https://doi.org/https://doi.org/10.1016/j.jad.2019.11.139

Scott, L. N., Levy, K. N., & Granger, D. A. (2013). Biobehavioral reactivity to social evaluative stress in women with borderline personality disorder. *Personality Disorders-Theory Research and Treatment, 4*(2), 91-100. https://doi.org/10.1037/a0030117

Shai, D., & Belsky, J. (2017). Parental embodied mentalizing: how the nonverbal dance between parents and infants predicts children's socio-emotional functioning. *Attachment & Human Development, 19*(2), 191-219. https://doi.org/10.1080/14616734.2016.1255653

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

Sibley, C. G., Fischer, R., & Liu, J. H. (2005). Reliability and validity of the revised experiences in close relationships (ECR-R) self-report measure of adult romantic attachment. *Personality and Social Psychology Bulletin, 31*(11), 1524-1536. https://doi.org/10.1177/0146167205276865

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

Sowden, S., & Shah, P. (2014). Self-other control: a candidate mechanism for social cognitive function [Opinion]. *Frontiers in Human Neuroscience, 8*(789). https://doi.org/10.3389/fnhum.2014.00789

Staebler, K., Renneberg, B., Stopsack, M., Fiedler, P., Weiler, M., & Roepke, S. (2011). Facial emotional expression in reaction to social exclusion in borderline personality

disorder. *Psychological Medicine, 41*(9), 1929-1938. https://doi.org/10.1017/S0033291711000080

Thomas, K. A., DeScioli, P., & Pinker, S. (2018). Common knowledge, coordination, and the logic of self-conscious emotions. *Evolution and Human Behavior, 39*, 179-190. https://doi.org/10.1016/j.evolhumbehav.2017.12.001

Thombs, B. D., Bernstein, D. P., Lobbestael, J., & Arntz, A. (2009). A validation study of the Dutch Childhood Trauma Questionnaire-Short Form: Factor structure, reliability, and known-groups validity. *Child Abuse & Neglect, 33*(8), 518-523. https://doi.org/10.1016/j.chiabu.2009.03.001

Verschuere, B., & Tibboel, H. (2011). The Dutch version of the McLean Screening Instrument for borderline personality disorder (MSI-BPD). *Psychologie & Gezondheid, 39*(4), 243-248. https://doi.org/DOI 10.1007/s12483-011-0046-0

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., & Sucharyna, T. A. (2013). Potential negative effects of perspective-taking efforts in the context of close relationships: increased bias and reduced satisfaction. *Journal of Personality and Social Psychology, 104*(1), 70-86. https://doi.org/10.1037/a0030184

Wang, L., & Maxwell, S. E. (2015). On Disaggregating Between-Person and Within-Person Effects With Longitudinal Data Using Multilevel Models. *Psychological Methods, 20*(1), 63-83. https://doi.org/10.1037/met0000030

Wood, A., Rychlowska, M., Korb, S., & Niedenthal, P. (2016). Fashioning the Face: Sensorimotor Simulation Contributes to Facial Expression Recognition. *Trends in Cognitive Sciences, 20*(3), 227-240. https://doi.org/https://doi.org/10.1016/j.tics.2015.12.010

Wu, H., Liu, X., Hagan, C. C., & Mobbs, D. (2020). Mentalizing during social InterAction: A four component model. *Cortex, 126*, 242-252. https://doi.org/https://doi.org/10.1016/j.cortex.2019.12.031

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., Vujanovic, A. A., Parachini, E. A., Boulanger, J. L., Frankenburg, F. R., & Hennen, J. (2003). A screening measure for BPD: The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). *Journal of Personality Disorders, 17*(6), 568-573. https://doi.org/DOI 10.1521/pedi.17.6.568.25355

Table 1

Mean scores per group and significance of group differences on all measures.

	High BPD (*n* = 30)			Low BPD (*n* = 31)			
	Mean	SD	Min; Max	Mean	SD	Min; Max	*p*
Age	18.40	0.62	18; 20	18.23	0.56	17; 20	0.25
MSI-BPD	7.67	0.96	7; 10	0.39	0.49	0; 1	< 0.001
Attachment Anxiety	4.32	0.99	2.2; 5.7	2.98	0.74	1.9; 4.4	< 0.001
Attachment Avoidance	3.81	1.05	1.8; 5.8	2.75	0.99	1.1; 5.2	< 0.001
Uncertainty about MS	0.96	0.48	0; 2	0.52	0.37	0; 1.3	< 0.001
Certainty about MS	0.54	0.46	0; 1.83	0.96	0.66	0; 2.7	0.006
Childhood Trauma	36.30	7.15	26; 56	29.74	3.91	25; 39	< 0.001

Note: MSI-BPD = Maclean Screening Inventory for Borderline Personality Disorder, MS = mental states.

Table 2

MLM predicting objective transparency based on BPD group status, and within- and between person centered estimated transparency and self-reported arousal, and their interactions.

Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.33	0.28 – 0.38	**<0.001**
Group	0.01	-0.06 – 0.09	0.746
Estimated Transparency (bpc)	0.04	0.01 – 0.08	**0.016**
Estimated Transparency (wpc)	0.01	-0.00 – 0.02	0.154
Group x Estimated Transparency (bpc)	-0.05	-0.10 – -0.01	**0.020**
Group x Estimated Transparency (wpc)	-0.00	-0.02 – 0.02	0.969
Subjective Arousal (bpc)	-0.05	-0.10 – 0.00	0.056
Subjective Arousal (wpc)	0.02	0.00 – 0.04	**0.043**
Group x Subjective Arousal (bpc)	0.04	-0.03 – 0.11	0.310
Group x Subjective Arousal (wpc)	-0.02	-0.05 – -0.00	**0.047**
Estimated Transparency (wpc) x Subjective Arousal (bpc)	-0.00	-0.01 – 0.01	0.889
Estimated Transparency (bpc) x Subjective Arousal (wpc)	0.02	0.01 – 0.03	**0.004**
Group x Estimated Transparency (wpc) x Subjective Arousal (bpc)	0.02	0.00 – 0.03	**0.039**
Group x Estimated Transparency (bpc) x Subjective Arousal (wpc)	-0.01	-0.02 – 0.00	0.140
Random effects			
ICC	0.67		
N_{ID}	61 $_{ID}$		
Observations	299		
Marginal R^2 / Conditional R^2	0.151 / 0.723		

Note: bpc = between-person centered; wpc = within-person centered; CI = confidence interval; ICC = Intraclass Correlation; N_{ID} = number of unique participants.

Table 3

Pearson correlation coefficients of self-report measures and transparency overestimation, and general transparency estimation error, averaged across the five clips.

	Transparency overestimation zET_min_zOT			Transparency estimation errors | zET_min_zOT |		
	Total	low-BPD	high-BPD	Total	low-BPD	high-BPD
Self-reported arousal	0.21	0.27	0.16	-0.09	0.20	-0.21
BPD features	-0.10	-0.27	-0.00	0.36**	0.10	0.24
Attachment Anxiety	-0.16	-0.07	-0.17	0.51 ***	0.14	0.56***
Attachment Avoidance	-0.17	-0.38*	-0.01	0.14	-0.10	0.02
Uncertainty about mental states	-0.15	-0.23	-0.07	0.29 *	0.06	0.21
Certainty about mental states	0.26 *	0.29	0.25	- 0.22	-0.02	-0.22
Childhood Trauma	-0.02	-0.36*	0.16	0.22	-0.07	0.11

Note: zET = standardized estimated transparency score; zOT = standardized objective transparency score; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

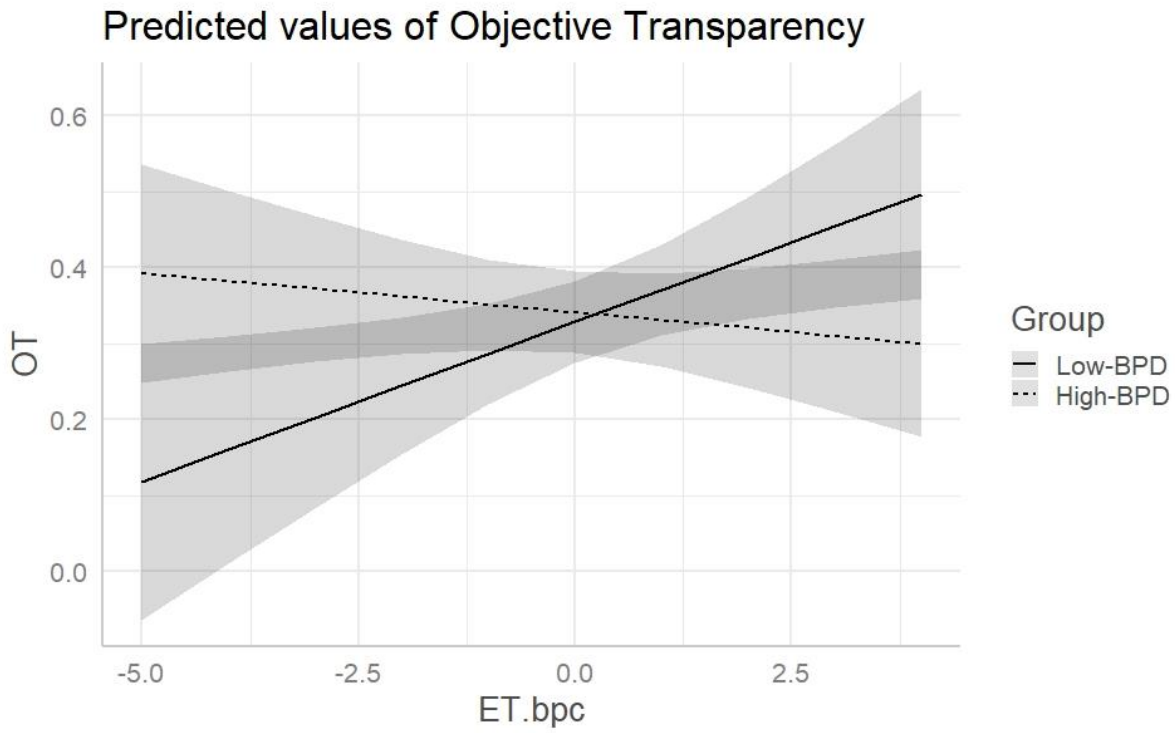

Figure 1. Between-person differences in estimated transparency (ET.bpc) predict objective transparency (OT) in the low-BPD group (full line) but not in the high-BPD group (dotted line).

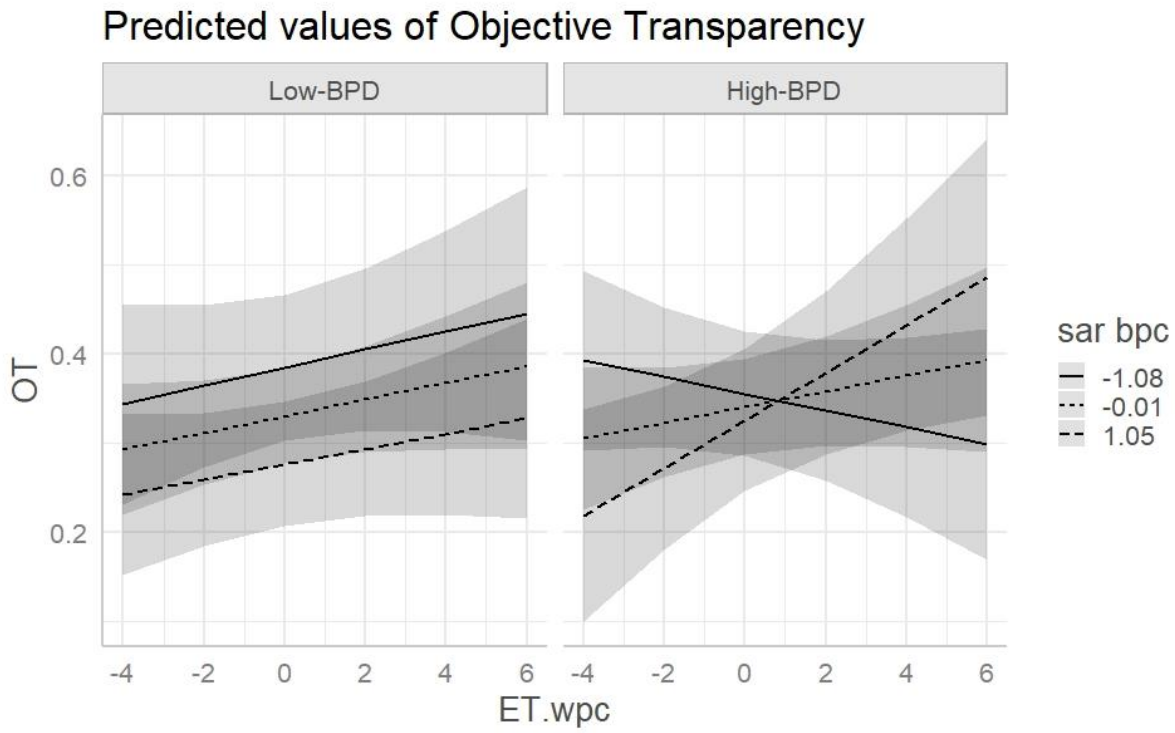

Figure 2. Within-person differences in estimated transparency (ET.wpc) predict objective transparency (OT) better in individuals in the high-BPD group who report overall high emotional arousal (SAR.bpc) (right panel, dashed line) and worse in individuals in the high-BPD group who report overall low arousal (right panel, full line), as compared to the low-BPD group (left panel).

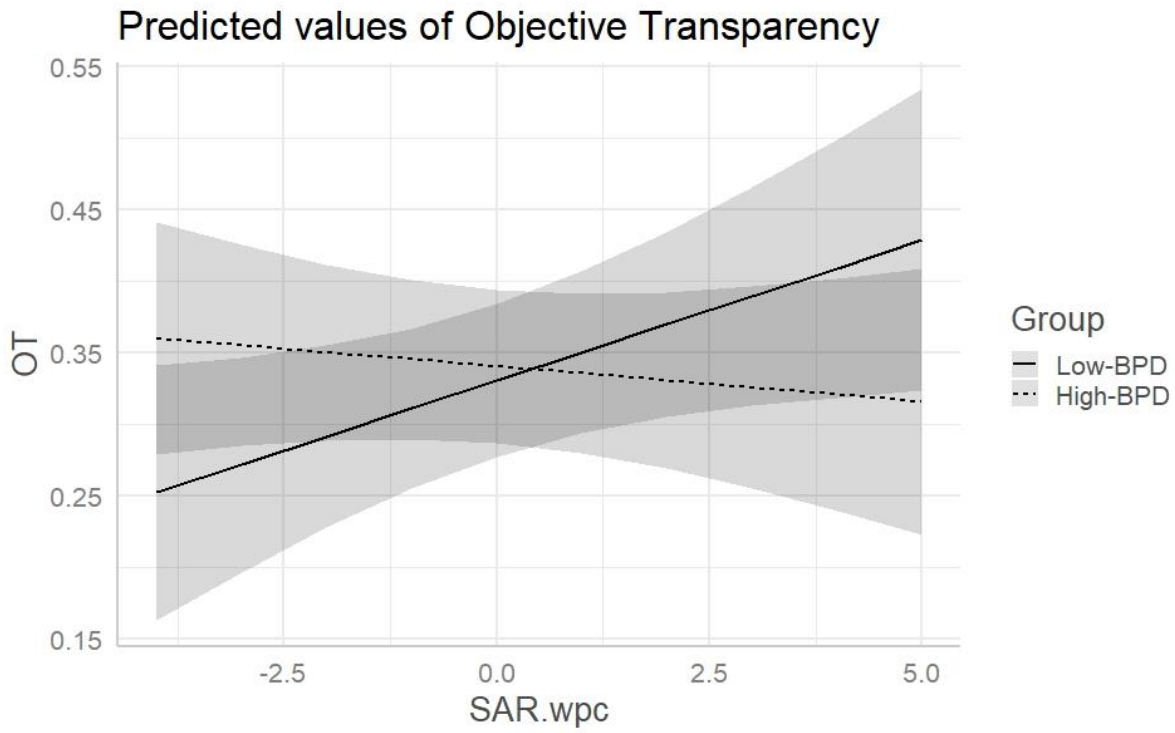

Figure 3. Within-person differences in emotional arousal (SAR.wpc) predict objective transparency (OT) in the low-BPD (red/dark grey line), but not in the high-BPD group (blue/light grey line).