Trustworthiness appraisal deficits in borderline personality disorder are associated with prefrontal cortex, not amygdala, impairment

Eric A. Fertuck, Jack Grinband, J. John Mann, Joy Hirsch, Kevin Ochsner, Paul Pilkonis, Jeff Erbe, Barbara Stanley

PII: S2213-1582(18)30363-2
DOI: https://doi.org/10.1016/j.nicl.2018.101616
Article Number: 101616
Reference: YNICL 101616
To appear in: NeuroImage: Clinical
Received date: 9 March 2018
Revised date: 5 November 2018
Accepted date: 25 November 2018

Please cite this article as: Eric A. Fertuck, Jack Grinband, J. John Mann, Joy Hirsch, Kevin Ochsner, Paul Pilkonis, Jeff Erbe, Barbara Stanley, Trustworthiness appraisal deficits in borderline personality disorder are associated with prefrontal cortex, not amygdala, impairment. Ynicl (2018), https://doi.org/10.1016/j.nicl.2018.101616

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Trustworthiness Appraisal Deficits in Borderline Personality Disorder are Associated with Prefrontal Cortex, not Amygdala, Impairment

Eric A. Fertuck$^{1,2,*}$ efertuck@ccny.cuny.edu, Jack Grinband$^{2,3,4,*}$ jackgrinband@gmail.com, J. John Mann$^{2,3}$, Joy Hirsch$^{4,5}$, Kevin Ochsner$^6$, Paul Pilkonis$^7$, Jeff Erbe$^{1,3}$, and Barbara Stanley$^{2,3}$

$^1$The City College of the City University of New York, Department of Psychology, NY, USA
$^2$Department of Psychiatry, Columbia University, New York, USA
$^3$New York State Psychiatric Institute, New York, NY, USA
$^4$Department of Radiology, Columbia University, New York, NY, USA
$^5$Department of Psychiatry and of Neurobiology, Yale School of Medicine, CT, USA
$^6$Department of Psychology, Columbia University, New York, NY, USA
$^7$Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

*Correspondence to: Eric A. Fertuck, Ph.D., The City College of New York, The City University of New York, NAC Room 7/239, 160 Convent Ave., New York, NY.  
*Correspondence to: Jack Grinband, Ph.D., Columbia University, 710 W. 168th St, Neurological Institute B04B, New York, NY 10032.

Abstract

Background

Borderline Personality Disorder (BPD) is associated with sensitivity to signals of interpersonal threats and misplaced trust in others. The amygdala, an integral part of the threat evaluation and response network, responds to both fear- and trust-related stimuli in non-clinical samples, and is more sensitive to emotional stimuli in BPD compared to controls. However, it is unknown whether the amygdalar response can account for deficits of trust and elevated sensitivity to interpersonal threat in BPD.
Methods
Facial stimuli were presented to 16 medication-free women with BPD and 17 demographically-matched healthy controls (total n=33). Participants appraised fearfulness or trustworthiness of the stimuli while BOLD fMRI was obtained.

Results
Though BPD participants judged stimuli as less trustworthy compared to controls, trustworthiness did not correlate with amygdalar activity in either group. Trustworthiness correlated with prefrontal regional activity in the insula and lateral prefrontal cortex. Prefrontal BOLD activity while appraising trustworthiness was smaller in BPD compared to controls, and the size of the reduction was proportional to each participant’s response bias.

Conclusions
Neural substrates of trustworthiness appraisal are associated with the lateral prefrontal cortex and insula, not amygdala, suggesting that untrustworthy stimuli do not elicit a subcortical threat response. Current models of BPD and its treatment may need to include a focus on improving impairments in frontally mediated trustworthiness appraisal in addition to amygdala-driven emotional hyper-reactivity.

Keywords: Borderline Personality Disorder, social cognition, fear, trust, emotion recognition, psychophysics, amygdala, prefrontal cortex, mentalization, interpersonal threat
1. Introduction

Heightened sensitivity to threat signals in interpersonal relationships and a misplaced trust in others are common vulnerabilities in Borderline Personality Disorder (BPD) (1-3). Individuals with BPD are prone to judge others as more hostile (4), are more likely to detect anger in ambiguous faces (5), to recognize angry faces faster than healthy controls (6), and to exhibit an elevated affective startle reflex (7). BPD is also associated with greater mistrust of others, characterized by a response bias during trustworthiness appraisal (8, 9). Furthermore, the emotional valence of a neutral face, i.e., the degree to which the face appears to be happy or angry, influences the visual assessment of trustworthiness in non-clinical individuals and has led to the hypothesis that appraisal of trustworthiness is actually an assessment of interpersonal threat (10). Thus, greater sensitivity to cues of interpersonal threat in BPD (11-13) may explain its association with elevated mistrust of others (8).

The neural mechanisms of threat appraisal have been studied extensively, and, it is widely accepted that the amygdala is an integral part of the threat appraisal and response system (14-16). The amygdala has also been proposed to be an important structure in the appraisal of trustworthiness (17-20). Bilateral lesions of the amygdala result in appraisals of elevated trustworthiness and approachability in both monkeys (21) and humans (17). Faces judged to be untrustworthy are associated with greater amygdala activity than trustworthy faces (18-20). Furthermore, after interpersonal betrayal, nasally administered oxytocin reduces amygdala activity, and preserves trust and cooperation (22). These findings suggest that, in non-clinical adults, appraisal of trustworthiness involves the amygdala, and cues of interpersonal threat, such as expressions of anger or
aggression, lead to an amygdala-based threat signal. By extension, greater mistrust of others in BPD may plausibly be a consequence of amygdala hyperactivity (8). In fact, several studies have reported that individuals with BPD exhibit greater amygdala activation to a wide range of interpersonal and emotional stimuli compared to controls (23-26), though hypoactivation has also been reported (27). However, though BPD has been associated with elevated amygdala activity to emotional stimuli and reduced interpersonal trust, a direct link between the elevated amygdala activity and impairment in trustworthiness appraisal has not been established. In the present study, facial expressions were systematically varied along the fearfulness or trustworthiness dimensions, and appraised by a BPD and a healthy control group. We tested the hypothesis that the response bias toward judging faces as untrustworthy, characteristic of BPD, will be correlated with amygdala hyperactivity. We also performed whole-brain analyses to determine whether other regions were related to trustworthiness appraisal deficits in BPD.

2. Methods and Materials

2.1 Participant Characteristics. All participants were female between the ages of 18 and 45 years; 17 were healthy controls and 16 had a DSM-IV diagnosis of BPD (28). Participants were recruited via advertisements and referral through a large, metropolitan hospital as part of ongoing clinical studies in mood disorders, suicidal behavior, and BPD. None of those with BPD were taking psychotropic medications while participating in the study, though sixty percent had a history of use of psychiatric medication. Exclusion criteria for the BPD group included a current major depressive episode, psychotic disorder, current substance use disorder, or a recent suicide attempt (in the last
6 months). The healthy control group was matched on demographics (age, ethnic/racial frequency, marital status), education level, and verbal IQ (the vocabulary subtest of the Wechsler Adult Intelligence Scale)(29), and was assessed with semi-structured interview to rule out a history of psychiatric or substance use disorder. Institutional Review Boards at two institutions approved the study. Fifty-eight participants signed consent, and 43 completed all assessments and the fMRI scan. Table 1 summarizes the demographic and clinical descriptions, and Supplementary Table 1 summarizes the clinical diagnoses of the BPD sample. Notably, 37.5% of the BPD group reported past of substance abuse or dependence, 68.8% had a past major depressive disorder, and none had a current or past bipolar or PTSD diagnosis.

2.2 Clinical Assessment. For individuals with BPD and controls, diagnoses were determined by Structured Clinical Interview for DSM-IV, Patient Edition (SCID-I) (Spitzer et al, 1990) and the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II) (30). Reliability studies within our research division yielded the following intra-class correlation coefficients (ICCs) (criterion levels are shown in parentheses): Axis I diagnosis/SCID-I, ICC=0.80 (0.70); Axis II diagnosis/SCID-II, ICC=0.70 (0.70); BPD diagnosis, ICC=0.89 (0.70). Depression severity was assessed using the Hamilton Depression Rating Scale (Ham-D; 31). Concurrent negative emotional state was assessed with the Profile of Mood States (32) a 65-item self-report questionnaire that provides a total score of state negative emotion scores based on 6 transient emotional states: tension-anxiety, depression-dejection, anger-hostility, confusion-bewilderment, vigor-activity, and fatigue. Hostility and aggression were assessed using the Buss Durkee Hostility Inventory (BDHI; 33). Abuse history was assessed as part of the demographic
interview, which asks participants whether they have experienced physical and sexual abuse before age 18. We assessed the number of prior suicide attempts from the Columbia Suicide History interview (34). The Rejection Sensitivity Questionnaire was used to assess anxious anticipation and expectation of interpersonal rejection (35) (See Table 1 for clinical characteristics).

2.3 Trustworthiness-Fear Face Appraisal Task. We utilized a task developed and validated by our group (8, 36) to measure an individual’s capacity to make subtle discriminations between facial features that indicate potential interpersonal threats, expressions of fear and of trustworthiness. Trustworthy faces were male, computer generated avatars selected from the stimuli developed and psychometrically validated by Todorov and colleagues (37). Facial fear stimuli were selected from the NimStim Face database (38) and identical to those used in Fertuck et al (8).

Faces at opposite extremes (neutral vs. fearful) or (trustworthy vs. untrustworthy) were morphed together in steps of 10% to create intermediate fear and trust values (Morpher software for Windows, version 3.1, M. Fujimiya). Individuals were presented with faces that varied along the fear or trustworthiness dimensions and asked to judge each face on a five-point Likert scale (where 1 is neutral or trustworthy and 5 is fearful or untrustworthy). (See (8) for more details on the development of the task and Supplementary Figure 1 and Supplementary Methods for sample stimuli and further elaboration of the procedure).

Subjective appraisal parameters were determined by fitting the behavioral data (i.e. rating versus % morph) to a logistic function of the form, \( y = \alpha + \beta/(1 + e^{\lambda(x+50)}) \) where \( x \) is the morph percentage of the stimulus, \( y \) is the mean subjective rating, and the
free parameters are $\alpha$ (the offset or bias), $\beta$ (the scaling or sensitivity), and $\lambda$ (the slope or discriminability) of the psychometric function. Each participant’s responses were checked to confirm that they completed the tasks as instructed (i.e. that subjective responses were not random but showed a monotonically increasing relationship with morph value). From those participants who completed the fMRI task, 2 BPD participants were excluded due to corruption of the data, and 4 BPD and 4 control participants were excluded because their ratings of either the trust or fear stimuli indicated a lack of discrimination between the most and least untrustworthy or fearful stimuli. All results, then, were based on data from 16 BPD patients and 17 healthy controls.

2.4 Functional Imaging

2.4.1 fMRI Parameters. Functional MRI was performed on a 1.5 Tesla GE Signa scanner using the EPI-BOLD sequence (TR = 2.0, TE = 86, flip angle = 34, number of slices = 27, array size = 64 x 64, voxel size = 3.1 mm x 3.1 mm x 4.0 mm, number of volumes = 150, duration of run = 6 min. Structural scans were performed using the 3D SPGR sequence (124 slices, 256 x 256, FOV = 200 mm).

2.4.2 fMRI Data Analysis. All analysis was done using the FMRIB Software Library (FSL 5.0.10; ([39]) and Matlab 2017a. Preprocessing consisted of motion correction (McFlirt), slice timing correction, high-pass filtering (> 50 sec), and spatial filtering (FWHM = 5 mm). Relative head motion of 0.5mm was set as a threshold and runs exceeding this value were excluded (none reached the threshold). Motion parameters (3 translations, 3 rotations, derivative and quadratic terms; 18 regressors total), CSF and white matter activity were included as confound regressors. Standard statistical parametric mapping techniques (FEAT) were performed in original T2* space. Group
analyses were performed using FEAT in MNI152 space at 2 mm isotropic resolution. Voxel-wise activation thresholds were set at $p = 0.05$, correction for multiple comparisons was done using Gaussian Random Field Theory with a cluster threshold of $p = 0.001$. A whole brain mask was used to exclude voxels outside the brain.

For each functional run, a regression model was created assuming three neural processes: (1) an unmodulated process, (2) the subjective appraisal of the stimulus, and (3) the quadratic term of the appraisal. The unmodulated regressor consisted of a set of boxcars in which each boxcar began at stimulus onset and ended when the subject made a response. The height of each boxcar was equal to 1 and represented any task-general activity (e.g. working memory, spatial attention, sensory processing, and other processes) that do not differ between conditions. The appraisal regressor had an identical temporal structure to the unmodulated regressor but the height of each boxcar was proportional to the participant’s subjective mean rating of the stimulus for the trust or fear decision. The quadratic regressor used an identical temporal structure to the appraisal regressor but with amplitude generated by demeaning the subject’s ratings and taking the absolute value. Trials with response times greater than 2.5 standard deviations outside the mean were excluded from the behavioral and imaging analyses. Each regressor was convolved with a custom HRF, which was individually estimated for each participant from their primary visual activity (40); custom HRFs have been shown to reduce both model error (41) and bias (42) relative to the canonical HRF. A fixed effects ($2^{nd}$ level, within subject) and a mixed-effects ($3^{rd}$ level, between subjects) analysis was done to compare patients with controls for the trust and fear appraisal regressors. We performed two ROI analyses of the amygdala. First, we created a mask by searching the Neurosynth database (43) using
the keyword “threat”. The reverse inference map was thresholded at 7 and binarized resulting in a bilateral amygdala mask positioned primarily over the lateral nuclei of the amygdala (MNI: -22,-2,-20; 24,-4,-20). A second analysis was performed subject-specific masks of threat-sensitive voxels. These voxels were identified as voxels modulated by subjective appraisal of fearfulness, thresholded at $>1.6$ and intersected with a whole amygdala mask. Both the Neurosynth mask and the subject-specific mask were used to average the parameter estimates of the masked voxels during trustworthiness appraisal. The Kolmogorov-Smirnov Test was used test for deviations from Normality for all t-tests (Supplementary Results). Cohen’s D ($d$) was computed as the group mean divided by sample standard deviation.

2.4.3 Assumptions. Our goal was to determine whether subjective appraisal of trustworthiness depends on threat signals generated by the amygdala. We assumed that fearfulness appraisal elicits threat signals in the amygdala and that any activity in the amygdala that increased with untrustworthiness would also represent a threat signal. Given these assumptions, if our paradigm could generate threat signals in the amygdala using fearful stimuli, it should also be able to generate amygdala threat signals using untrustworthy stimuli. Furthermore, since BPD is associated with elevated sensitivity to social threat and a bias toward judging others as untrustworthy, BPD subjects should show elevated threat activity in the amygdala compared to controls using untrustworthy stimuli.

3. Results
Consistent with our previous study (8), the BPD group showed a response bias to judge faces as untrustworthy (t-test of bias: control, M = 1.6, SD = 0.15; BPD, M = 2.1, SD = 0.16, t(28) = 2.44, z=3.23, p = 0.02) and had a smaller dynamic range, or, sensitivity (t-test for scale: control, M = 3.08, SD = 0.41; BPD, M = 1.71, SD = 0.23, t(28) = 2.8, z=3.98, p < 0.01). Trustworthiness appraisal did not result in significant group differences in discriminability (Figure 1). Appraisal of fearfulness did not show any significant group differences for bias (p = 0.47), sensitivity (p = 0.14), or discriminability (p = 0.49). An analysis of variance showed that the BPD group exhibited longer RTs than controls (Figure 1) for trustworthiness (rating, p = 0.002; group, p < 0.0001) and fearfulness (rating, p < 1x10^{-6}; group, p = 0.007), and no significant interactions.

To identify the neural structures associated with the two types of appraisals, we performed a whole brain analysis, regressing the appraisal ratings made by each subject on the BOLD data. Consistent with most fMRI studies of fear processing (44, 45), both amygdalae were robustly modulated by subjective appraisals of the fearful stimuli – BOLD magnitude increased as a function of the subjective rating of intensity of the stimulus (Fig 2A; peak response, MNI: 24, -8, -14, Z = 3.83; -24, -4, -14, Z = 3.56; Supplementary Table 2). If the threat-related cues detected by the amygdala are also important for trustworthiness appraisal, then amygdala activity should be modulated by trustworthiness. However, the whole brain analysis showed no activity in the amygdala that was significantly modulated by stimulus trustworthiness (Fig 2B; Supplementary Table 2).
Averaging across voxels can improve the signal to noise ratio; thus, we performed an ROI analysis of the amygdala using a mask generated on Neurosynth using the keyword “threat”. In healthy controls, the mean activity of voxels within the mask showed robust amygdala modulation by subjective fear ratings ($p = 0.01$, $d = 0.70$) but contrary to previous work (18, 20), no significant modulation by subjective trust ratings ($p = 0.33$, $d = 0.25$). In the BPD group, no significant activity was detected either by fear ($p = 0.33$, $d = 0.25$) or trust ($p = 0.56$, $d = -0.15$) ratings. Since BPD is associated with elevated sensitivity to interpersonal threats (4, 46), if the amygdala were sensitive to untrustworthiness, then BPD subjects should show greater amygdala activation than controls as the stimuli become less trustworthy. However, a comparison of the two groups showed no significant difference between groups for trust ($p = 0.26$, $d = 0.40$) or fear ($p = 0.33$, $d = 0.35$).

Previous studies have suggested that the effect of trustworthiness appraisal on amygdala activity in healthy controls is best described by a quadratic relationship (19) (47, 48). Though our controls showed a significant quadratic relationship for fear ($p = 0.04$, $d = 0.53$), no significant quadratic relationship for trust ($p = 0.80$, $d = 0.06$) was found. In the BPD group, the quadratic model was not significant for fear ($p = 0.68$, $d = 0.18$), but was significant for trust ($p = 0.02$, $d = 0.65$).

To determine whether the two tasks activate similar brain networks, we compared whole-brain activations (Fig 4A). Fearfulness appraisal activated primarily sub-cortical regions, whereas trustworthiness appraisal was associated primarily with cortical activity. To dissociate activity specific to fearfulness and trustworthiness appraisal from general decision-making activity related to stimulus intensity, we performed a contrast between
task conditions (contrasting fearfulness > trustworthiness and trustworthiness > fearfulness on the appraisal regressor; Fig 4B). Using a cluster threshold of $p = 0.001$, fearfulness > trustworthiness did not result in significant activations. However, because the amygdala nuclei are small structures, $p = 0.001$ may result in elevated Type II error in subcortical structures. At a cluster threshold of $p = 0.05$, fearfulness-specific activity was localized to subcortical regions, i.e., amygdala and ventral striatum (peak response, MNI: 22, -6, -8), consistent with the previous ROI analysis (i.e. Fig 2). Moreover, even at a more liberal threshold, no fearfulness-specific activity in the cortex was detected. In contrast, trustworthiness-specific activity was present only in cortical regions, broadly distributed across posterior parietal cortex, and dorsolateral and mediolateral prefrontal cortex, and no spatial overlap of amygdala (Supplementary Table 3).

Because BPD is associated with behavioral abnormalities in trustworthiness appraisal, we hypothesized that the trustworthiness-specific network (i.e. trustworthiness > fearfulness) would show activity differences between BPD and control subjects. We, thus, performed the following contrast: (trustworthiness > fearfulness)$_{\text{Control}} >$ (trustworthiness > fearfulness)$_{\text{BPD}}$. BPD participants had lower trustworthiness-specific activity in prefrontal cortex (Fig 5A), especially anterior insula and lateral PFC (Supplementary Table 3). Finally, to determine whether these group differences were related to individual subjects’ decision variables, we intersected voxels that showed activity specific for trustworthiness appraisal (Fig 4B, trustworthiness > fearfulness) with voxels that differed between groups (5A, (trustworthiness > fearfulness)$_{\text{Control}} >$ (trustworthiness > fearfulness)$_{\text{BPD}}$) and compared them to individual differences in response bias and sensitivity. The anterior insula and lateral PFC (Fig 5B) activity was
related to the degree of bias \( (r = 0.457, p = 0.007) \) and sensitivity \( (r = 0.597, p = 0.0005) \) impairment in trustworthiness appraisal (Fig 5C), such that, the weaker the network activity, the greater the bias toward untrustworthy ratings and the smaller the range of responses.

4. Discussion

The amygdala is an integral part of the threat detection system in humans (14-16), and to the extent that untrustworthy faces represent interpersonal threats, investigators have argued that the amygdala is integral to the appraisal of trustworthiness in non-clinical adults (18-20). Furthermore, individuals with BPD have been shown to have response biases toward mistrusting others (8, 9, 49) and hyperactive responses of the amygdala to emotional stimuli (23-26). Our goal was to test whether amygdala hyperactivity could explain the response biases in BPD during the appraisal of trustworthiness (8, 9, 49). Surprisingly, we found no relationship between trustworthiness appraisal and amygdala activity, and no difference in amygdala activity between BPD and control participants. Instead, trustworthiness appraisal deficits in BPD were associated with blunted prefrontal activity in anterior insula and lateral PFC compared to controls.

Evidence that trustworthiness activates the amygdala has been inconsistent. Studies that categorically compared trustworthy versus untrustworthy stimuli typically find greater amygdala responses to untrustworthy faces (50-53). Similarly, some parametric studies have demonstrated that amygdala activity increases monotonically with untrustworthiness (18, 20). However, others found a quadratic, not monotonic,
relationship, between trustworthiness and amygdala responses (47, 48). Contrary to these previous studies, we found no evidence that amygdala activity increases monotonically or quadratically with untrustworthiness in healthy controls. This lack of response was not due to sensitivity of our behavioral paradigm. In fact, consistent with our previous studies (8, 9, 49), our behavioral data showed a sigmoidal relationship between stimulus and response, and a response bias in BPD for judging stimuli as less trustworthy, but not more fearful. Moreover, the trustworthiness-stimuli were psychometrically discriminable by both groups with a dynamic range similar to the fearful stimuli and the fearful stimuli elicited robust, bilateral amygdala responses that scaled parametrically with subjective intensity. This suggests that if trustworthiness decisions depended on threat-related amygdala activity, modulation of amygdala by trustworthiness would have been detectable with our paradigm.

Previous parametric studies focused mostly on “implicit,” or sub-conscious, processing of trustworthiness, distracting subjects from the trustworthiness dimension with an irrelevant task (18, 20, 47) or using very short (200 ms) stimulus durations (48). While implicit trustworthiness processing is commonly referred to as “trustworthiness decisions,” it is not clear that any amygdala activity that is correlated with trustworthiness, but also lacks an associated behavioral response, actually represents a decision process. Instead, this activity is more likely to be related to low-level, perceptual processing (45, 54-59). In fact, trustworthiness has been shown to be decomposable into two perceptual factors – dominance and emotional valence, where emotional valence is expressed as facial features ranging from happy to angry (10, 60). However, while anger has been shown to represent a cue for untrustworthiness, a meta-
analysis of 105 imaging studies has not found it to reliably activate the amygdala (61). Moreover, because the amygdala generally responds to emotional faces (61), even at sub-threshold levels (45, 54-59), the implicit or rapid processing of trustworthiness by the amygdala may actually reflect the emotional valence detectable in the stimulus rather than the appraisal of trustworthiness per se.

Facial cues associated with low trustworthiness are not necessarily reliable or immediate expressions of threat, compared to reliable cues such as an image of a snake or a pointed gun. Rather, trustworthiness appraisal may be better conceptualized as a probabilistic prediction about the likelihood of interpersonal betrayal or exploitation by others. Probabilistic reasoning, especially in social contexts, has been associated with prefrontal cortical processing (62-64). Our results show that trustworthiness is mediated by prefrontal cortical (posterior parietal cortex, anterior insula, and lateral PFC) activity and that trustworthiness appraisal deficits in BPD are also mediated by the same regions.

The trustworthiness appraisal impairments identified here may help elucidate mechanisms of turbulent relationships in BPD. Individuals with BPD maintain unstable interpersonal ties, as they oscillate between establishing new relationships and ending them (65). Some of the most high risk diagnostic criteria of BPD such as self-injury, suicidality, intense and inappropriate anger, impulsivity, and heightened emotional sensitivity are mediated by the quality of interpersonal bonds between the person with BPD and significant others (66). Facial expressions within in interpersonal contexts are salient stimuli, and can anticipate mistrust and the expectation of rejection (9, 67-69). Consequently, the trustworthiness appraisal impairments in BPD can increase their propensity interpersonal conflicts, lead to uncooperative exchanges in social interactions,
threaten the formation of new relationships, and undermine long-term relationships. The trustworthiness discriminability impairment mediated by prefrontal cortex processes may help clinicians to understand commonly observed interpersonal dynamics in BPD. For instance, individuals with BPD often reflexively enter into new relations with questionable partners, while simultaneously expressing extreme caution and suspiciousness towards presumably helpful and supportive others.

Improving accurate appraisal of trustworthiness in interpersonal and therapeutic relationships in BPD may be crucial to therapeutic improvement, and dissociating the roles of prefrontal cortex and amygdala in trustworthiness appraisal may aid in sharpening intervention targets. Prominent, evidence-based therapies for BPD such as Transference Focused Psychotherapy (TFP, 70) and Mentalization-Based Therapy (MBT, 71), focus implicitly and explicitly (72) on enhancing trustworthiness appraisal by fostering frontally-mediated social reappraisal processes. However, there may yet be untapped strategies and interventions that those with BPD, such as improving accurate probabilistic reasoning around trustworthiness appraisals.

Limitations. Without a psychiatric control group, the specificity of the trustworthiness impairment findings has yet to be established. However, we have published work using the same trustworthiness and fear tasks in a PTSD sample compared to a trauma-exposed/no PTSD control group and a healthy control group. The PTSD group showed a response bias toward judging stimuli as more trustworthy compared to the trauma-exposed controls (36). This is opposite to our BPD findings, which show a bias toward less trustworthy appraisals, and suggests some clinical specificity of our results. Finally, although our BPD sample has relatively few co-
morbidities, the mean Global Assessment of Functioning (GAF) score of the group was 55.12, consistent with multi-site, longitudinal studies of BPD (73) and suggesting that our BPD group had comparable severity of illness.

**Conclusions.** In summary, we found no evidence of amygdala hyperactivity in BPD subjects during appraisal of trustworthiness. Our results show, however, that trustworthiness biases in BPD involve higher order prefrontal cortical regions. Additionally, further study is needed to clarify impact of emotional expressions (e.g., appraisal of anger in facial stimuli may overlap with untrustworthiness perception) on trustworthiness appraisal and amygdala activity in BPD and comparison groups.

**Acknowledgements**

Eric A. Fertuck and Jack Grinband contributed equally to this manuscript. This study was supported in part by grants from NIMH (MH077044) and the American Psychoanalytic Association to EF. Jamie Wilsnack and the staff of the Molecular Imaging and Neuropathology Division of New York State Psychiatric Institute were instrumental in the conduct of this study. The authors have no conflicts of interest.
References


Highlights

- Borderline Personality Disorder (BPD) is associated with sensitivity to signals of interpersonal betrayal and misplaced trust in others.

- The amygdala, an integral part of the threat evaluation and response network, responds to both fear- and trust-related stimuli in non-clinical samples, and is more sensitive to emotional stimuli in BPD compared to controls.

- However, in the present study, neural substrates of trustworthiness appraisal are associated with the lateral prefrontal cortex and insula, not amygdala.

- BPD exhibits impairments in prefrontal activity when judging trustworthiness compared to healthy controls.

- Current models of BPD may need to include a focus on improving impairments in frontally mediated trustworthiness appraisal in addition to amygdala-driven emotional hyper-reactivity.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5

(A) Controls vs. BPD brain images showing differences in activity levels.

(B) Another set of brain images with a different scale and parameter estimates.

(C) Graphs depicting bias and scale against parameter estimates for Controls and BPD groups.