Atypical Neural Plasticity and Behavioral Effects of Trustworthiness Learning in High Vs. Low Borderline Personality Disorder Features: An Experimental Approach

Eric Fertuck (efertuck@ccny.cuny.edu)
City College of New York https://orcid.org/0000-0003-3785-6630
Stephanie Fischer
The City College of New York
Robert Melara
The City College of New York

Research article

Keywords: Borderline Personality Disorder, social cognition, trust, neuroplasticity, electroencephalogram

Posted Date: October 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-87567/v1

Abstract

Background: The ability to accurately decide who is trustworthy, and to, in the face of new information, adjust judgment of others' trustworthiness accurately, flexibly, and efficiently is clinically impaired in Borderline Personality Disorder (BPD).

Methods: A novel trust learning paradigm was administered to high (H-BPD) and low (L-BPD) in BPD features undergraduate participants. Neutral faces were paired with trust-relevant behaviors in each of four conditions: trustworthy, untrustworthy, mixed, and ambiguous. After training, participants rated faces on trustworthiness as electroencephalographic measures were recorded.

Results: H-BPD rated neutral faces as significantly more untrustworthy than L-BPD at both time periods. Negative and ambiguous trustworthiness pairing conditions led to lower trustworthiness ratings, whereas trustworthy and mixed descriptors led to higher trustworthiness ratings. Training enhanced the amplitude of an early sensory ERP component (i.e., P1) for both groups. The slow wave ERP, an index of sustained attention, revealed greater focus after training to trustworthy descriptors in H-BPD and to untrustworthy descriptors in L-BPD.

Conclusions: Social learning can modify an untrustworthiness bias in BPD at neural and behavioral levels. The results suggest that differential neural plasticity may account for the negative trustworthiness appraisal bias in BPD, and that interventions targeting frontal, attentional processes during trustworthiness learning may be a key mechanism of therapeutic change.

Background

The ability to accurately decide who is trustworthy, and to, in the face of new information, adjust judgment of others' trustworthiness accurately, flexibly, and efficiently is critical for the navigation of unpredictable and dynamic human social systems. Individuals with Borderline Personality Disorder (BPD), in particular, exhibit extreme distress and confusion navigating such social systems. These individuals often experience other people as threatening, hostile, and rejecting (Barnow et al., 2009; Beck et al., 2001) and are biased to judge others as untrustworthy (Fertuck, Fischer, & Beeney, 2019; Fertuck, Grinband, & Stanley, 2013; Fertuck et al., 2019; Miano, Fertuck, Arntz, & Stanley, 2013). Clinically, these appraisals appear resistant to change or learning-based modification, obstructing progress in treatment and threatening the therapeutic alliance fundamental to benefit from psychotherapy (Fertuck, Fischer, & Beeney, 2019). The goal of the current study was to explore the feasibility of a new social learning paradigm that seeks to modify, through experience, appraisals of trustworthiness in undergraduate participants with high or low BPD symptoms. A neural plasticity approach can inform the mechanisms of such appraisals and their resistance to learning, thus highlighting potential mechanisms of change in treatment in BPD.

Impairments in trust appraisal in BPD

Untrusworthiness bias. Numerous studies have found that individuals with BPD tend to judge other people as untrustworthy (Fertuck et al., 2013; Fertuck, Grinband et al., 2019; Miano et al., 2013; Nicol et al., 2013). In one study using facial stimuli morphed to show different degrees of trustworthiness and fear, people with BPD rated faces as less trustworthy than healthy controls across varying degrees of trust, whereas their appraisals of fear did not significantly differ from controls (Fertuck et al., 2013). Moreover, their response times for evaluating ambiguously trustworthy faces were longer compared with controls, suggesting increased mental effort in evaluating trustworthiness. fMRI findings indicated impaired lateral prefrontal activity in BPD, not amygdala, is associated with trustworthiness impairment in BPD (Fertuck, Grinband et al., 2019); there were not difference between BPD and controls in fear processing. These results suggest that people with BPD have a response bias for evaluating others as less trustworthy that is associated with impaired prefrontal processing.

A possible theoretical basis of the untrustworthiness bias in BPD implicates rejection sensitivity as mediating the relationship between trustworthiness and BPD symptoms (Miano et al., 2013). People high in rejection sensitivity, including BPD, have an intense need for connection with others, partly due to a lack of confidence in their ability to cope with social stressors (Berenson et al., 2009; Romero-Canyas et al., 2010; Staebler et al., 2011; Watson & Nesdale, 2012). Concurrently, people with BPD also feel more easily excluded in social settings (Gutz, Renneberg, Roepke, & Niedeggen, 2015) and tend to react to rejection with anger or rage (Berenson et al., 2011). It is thus conceivable that getting hurt routinely by others, due to real or perceived rejection, conditions one to be especially cautious and mistrustful of others (Miano, Fertuck, Arntz, & Stanley, 2013).

Trust in neuroeconomic games. Trust biases in BPD also have been revealed in studies employing neuroeconomics games to explore social decision making (Franzen et al., 2011; Unoka et al., 2009). Making adaptive social exchanges often depends on the ability to know when to trust others and when to withdraw trust of others flexibly (Seres et al., 2009). In the Trust Game, two players – an investor and a trustee – work collaboratively to maximize monetary gain. The investor decides on each round how much to invest, keeping the difference. The invested amount triples and is held by the trustee, who then decides how much to return to the investor. The amount of money invested is tightly indexed by insula activation, which in healthy participants playing trustees is associated with lower activation for large investments. However, individuals with BPD report less trust for others in the Trust Game, repay less money to investors, and, as trustees, show no relationship between investment and insula activation. Moreover, as investors, people with BPD report less trust of trustees, are less optimistic about the amount of money that will be returned to them, and invest less in their partner relative to healthy controls or people with depression (King-Casas et al., 2008). Yet when investments are sent to a random lottery (non-social condition), people with BPD invest similarly to other groups, suggesting that mistrust in BPD is directed specifically toward social rewards. These results, indicate that individuals with BPD hold untrustworthy mental representations of people that foster mistrust and pessimism, a tendency absent in non-social exchanges.

Event Related Potential (ERP) correlates of BPD symptoms

Several neuroimaging studies reveal an early sensory basis for the social difficulties (Berchio et al., 2017; Merkl et al., 2010) and affective instability (Koenigsberg et al., 2014) in BPD. Event-related potentials (ERPs), used to illuminate different stages of information processing, have uncovered differences between BPD and healthy controls in early visual processing (e.g., P1 component), mid-latency facial processing (e.g., N170 component), and late-state attentional processing (e.g., P3 component) (e.g., Marissen, Meuleman, & Franken, 2010; Meares, Melkonian, Gordon, & Williams, 2005; Meares, Schore, & Melkonian, 2011; Ruchsow et al., 2008). Hidalgo et al. (2016) found that BPD participants demonstrated a negativity bias compared to healthy controls in the classification of happy faces, with enhanced P1 amplitude and reduced N170 and P300 amplitudes. Yet there is a paucity of information regarding the neural correlates of behavior or therapeutic change in BPD and no extant studies examining the neural basis of the mistrust bias. Thus, a central goal of the current investigation was to probe for group differences in ERP amplitudes to trustworthy faces before and after social learning.

Neuroplasticity from cognitive training and social learning

Recent research suggests that rigorous training to more efficiently employ certain executive functions, including inhibitory control and working memory, can yield long-lasting improvements to behavioral performance on cognitive tasks (Diamond & Lee, 2011; Jaeggi, Buschkuchi, Shah, & Jonides, 2013; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009), including relief from response bias (Schmidt, Richey, Buckner et al., 2009) and limited transfer to untrained tasks (Garner, Matthews, Remington & Dux, 2015). Cognitive training yields lasting changes in neural activity (Millner, Jaroszewski, Chamarthi, and Pizzagalli, 2012; Rueda, Checa, & Combita, 2008; Tang & Posner, 2009) – *neuroplasticity* – especially among those with poor cognitive abilities (Diamond & Lee, 2011). Melara, Singh, and Hien (2018) found improvement from cognitive training in the speed and efficiency of conflict resolution (see also Garner, Matthews, Remington & Dux, 2015), particularly in individuals with poor attention skills. The plasticity in executive control suggested by training results has implications for treating psychiatric syndromes involving compromise to executive functions of working memory, including individuals with BPD. Schmidt, Richey, Buckner et al. (2009), for example, trained participants with social anxiety disorder to avert threatening cues during performance of the dot-probe paradigm. The bias training led to a significant decrease in social phobia and anxiety scores relative to a placebo control, an improvement that was maintained over a 5-month follow-up. The therapeutic effect compared well with other modalities commonly used for treating social anxiety, including cognitive behavior therapy, exposure therapy, and psychopharmologic therapy.

Relatively few studies have explored social learning outcomes in individuals with BPD. Fineberg (2018) investigated the use of social and non-social cues in a learning task. BPD participants were able to make use of both types of cues, but at only half the learning rate of healthy controls, weighing social over non-social cues to make their decisions, with negative social experience (incorrect advice) having a less potent and less durable influence on choice than positive social experience (correct advice). In the current study, we exploit the power of social cues to explore whether the mistrust bias in BPD can be unlearned.

The current study

The present study extends previous research on the untrustworthiness bias to neutral faces in BPD (Fertuck, Grinband, & Stanley, 2013) by using a novel trust learning paradigm to shift (increase or decrease) trustworthiness appraisals in human faces in participants high (H-BPD) and low (L-BPD) in BPD symptoms. Four neutral faces (identities) were paired with trait-relevant behaviors, embodied in four types of social learning conditions: (1) positive trustworthiness traits (e.g., “this person helped an elderly woman across the street”), (2) negative trustworthiness traits (e.g., “this person spreads negative gossip about friends”), (3) a mixed combination of positive and negative trustworthiness traits, and (4) ambiguous trustworthiness traits (e.g., “this person ignored an old girlfriend at a party”). After training, participants were asked to rate the four identities on trustworthiness as electroencephalographic measures were recorded. We asked whether trustworthy (or, untrustworthy) paired traits increased (decreased) trustworthiness appraisals. We hypothesized that ambiguous traits would maintain the usual untrustworthiness response bias whereas integrating mixed trustworthy and untrustworthy traits might allow for the possibility of a positive response bias. We were especially interested in whether social learning has a measurable behavioral or neurophysiological influence on appraisals of individuals high in BPD symptomology, who are otherwise resistant to therapeutic change in trustworthiness appraisal accuracy.

Method

Participants. Twenty-five participants (15 females, average age = 20.4 years), recruited from an urban, diverse undergraduate community, were given Psychology course extra credit and compensation ($10/hour) for participation in a two-phase study. The nature of the procedures was explained fully and informed consent was obtained from each participant in each phase; the Institutional Review Board of the City University of New York approved the protocol. All participants were fluent in English, had normal or corrected-to-normal vision, and no history of neurological disorder (self report). Participants were screened and recruited into two groups, those high (H-BPD) versus low (L-BPD) in BPD features, based on responses on the top and bottom quartile to the Personality Assessment Inventory (PAI) – BPD Subscale (Morey & Zanarini, 2000) (see Procedure for details).

Stimuli and apparatus. Gray-scale face stimuli representing four different identities (hereafter, neutral faces) from the Todorov (2012) stimulus set were varied parametrically along the psychological trait of trust using facial morphing software (Fertuck et al, 2013; Fertuck et al., 2019; Todorov, Baron, & Oosterhof, 2008). Seven morphs were created for each of the four neutral faces, with each morph containing features that were either more or less trustworthy than the original neutral faces. For each participant, the four neutral stimuli were assigned randomly without replacement to one of four training conditions: (1) trustworthy, (2) untrustworthy, (3) ambiguously trustworthy, and (4) mixed trustworthiness. Face and trait stimuli were presented using Presentation software on a PC Windows computer.

Procedure. Participants were tested in two phases. In the first phase participants were screened using the Rosenberg Self-Esteem Scale (Rosenberg, 1965) and the PAI-BPD. The Rosenberg Self-Esteem (RSE) is a 10-item survey that measures global self-worth by assessing both positive and negative feelings about

the self. The PAI-BPD is a 24-item self-report inventory scale that focuses on attributes indicative of BPD. Subscales are: Affective Instability (poor regulation of emotional responses; BOR-A, 6 items), Identity Disturbances (uncertainties about sense of self and overall feeling of lack of purpose; BOR-I, 6 items), Negative Relationships (history of abusive, and intensely unstable relationships; BOR-N, 6 items), and Self Harm (self-destructive and impulsive behaviors without concern for negative outcomes; BOR-S, 6 items). Questionnaires were administered on individual computers in a computer laboratory using the Qualtrics platform. Participants whose composite score on the PAI-BPD fell into either the top or bottom quartiles were assigned to the H-BPD or L-BPD group, respectively, and invited to return to participate in the second phase of the study. For all other participants, the study terminated with the screening phase.

The second phase comprised an identity learning task, a trustworthiness rating task, and a trustworthiness learning task (see Fig. 1). In identity learning, participants were asked to associate one of four male names (Jay, Robert, Bill, and Michael were randomly paired with four facial identities) with each neutral face identity. Participants were continuously tested on the identity of the n-1 neutral face using a forced-choice task requiring them to select the facial identity that matched a paired name from the previous trial (“one back identity task”). Identity learning was complete when participants exceeded a criterion of 90% accuracy.

Participants then completed a rating task in which the four neutral stimuli and their corresponding morphs appeared one at a time in random order while participants rated each stimulus on trustworthiness using a Likert scale (1 = extremely trustworthy; 5 = extremely untrustworthy). Participants made each rating twice, for a total of 64 trials. The response time to make each rating was also recorded.

Participants then completed a social learning task in which each of the four neutral faces was paired with a trait descriptor (i.e., trustworthy, untrustworthy, ambiguous, and mixed). For example, in the untrustworthy condition, a face was associated with the descriptor, e.g., “spreads negative gossip about his friends.” In the trustworthy condition, a face was paired with items such as, “practices what he preaches.” In the ambiguous condition, a face was paired with phrase such as, “ignores his old girlfriend at parties.” Each condition contained two possible trait descriptors of a trustworthy or untrustworthy valence; the mixed condition combined one trustworthy trait with one untrustworthy trait. Facial identities were assigned randomly to these four conditions. Participants were required to select the facial identity that matched a paired trust related trait from the previous trial (“one back trait task”). Learning was assessed through forced-choice to 90% accuracy criterion. After completion of social learning, participants repeated the rating task. The entire experiment lasted approximately three hours.

Data recording and analysis. Ratings of the untrustworthiness of the faces, and rating reaction times (RTs), were averaged for each participant in each condition separately at pretest and posttest. We performed mixed model analyses of variance (ANOVAs) on behavioral data using Statistica® software, with Group (2 levels: low BPD, high BPD) as the between-subjects factor and Test (2 levels: pretest, posttest) and Condition (4 levels: ambiguous, untrustworthy, trustworthy, mixed) as within-subjects factors. To guard against violations of the sphericity assumption with repeated-measures data, all main

effects and interactions reported as significant were reliable after Greenhouse-Geisser correction (Greenhouse & Geisser, 1959).

Continuous recordings of the EEG were collected at a sampling rate of 512 Hz using ANT Neuro system in a high-density (128 electrodes) montage arranged in an elastic cap. Blinks and other eye movements were monitored by electrooculogram (EOG) from two electrode montages, one on the infra- and supra-orbital ridges of the right eye (VEOG), the other on the outer canthi of each eye (HEOG). Trials containing mastoid activity exceeding 100 μV were rejected. Trials contaminated by blinks, eye movements, or other movement artifacts were defined as z-values on the VEOG, HEOG, and lowermost scalp channels exceeding 4.5 in a frequency band between 1 and 140 Hz; artifact trials were removed automatically using a Matlab routine (Fieldtrip; Oostenveld, Fries, Maris, & Schoffelen, 2011).

Sweep time to each face stimulus was 1200 ms, including a 200 ms pre-stimulus baseline; signal-averaged waveforms were referenced to linked mastoids band-pass filtered between .1 and 30 Hz. Peak amplitude and latency were measured to four ERP components: P1, P2, N2 and late positive slow wave. P1 was measured over eight posterior electrode locations (LL13, L13, Z13, R13, RR13, L13, Z14, R14) during a search epoch 34–68 ms after stimulus onset. P2 (148–205 ms after stimulus onset) and N2 (182–250 ms) were measured over eight anterior electrode sites (LL3, LL6, Z2, Z3, Z4, Z5, Z6,f RR3). The late slow wave, a measure of sustained attention (Näätänen, 1992), was defined as the average voltage to the face stimulus 700–900 ms after stimulus onset and measured over eight central electrode locations: Z5, L5, R5, Z6, L6, R6, RR5, RR6 (Bidet-Caulet et al., 2010). ANOVAs of ERP amplitudes mimicked behavioral analyses, with Group as the between-subjects factor, and Test and Condition as within-subjects factors. All main effects and interactions reported as significant were reliable after Greenhouse-Geisser correction (Greenhouse & Geisser, 1959).

Results

Power Analysis. A statistical power analysis was performed for sample size evaluation, based on our data comparing H-BPD to L-BPD groups. With an alpha = .05 and power = 0.80, the projected sample size needed do detect “medium” effect size differences between groups (GPower 3.1) is approximately N = 12 for this between group comparison. Thus, our sample size is adequate for the main hypotheses of this study.

Behavioral analyses. A mixed model ANOVA was performed on ratings of untrustworthiness, with Group (2 levels: low BPD, high BPD) as the between-subjects factor and Test (2 levels: pretest, posttest) and Condition (4 levels: ambiguous, negative, positive, mixed) as within-subjects factors. We found a main effect of Group, $F(1,25) = 8.32$, $p < .01$, $MS_e = .53$, $\eta^2 = .44$; participants high in BPD features judged faces significantly higher in untrustworthiness (average = 3.23/5.0) than participants low in BPD features (2.94/5.0). There were no main effects of Test, $F(1,25) = 1.91$, *ns*, $MS_e = .09$, $\eta^2 = .02$, or Condition, $F(3,75) = 1.40$, *ns*, $MS_e = .60$, $\eta^2 = .25$, but there was a significant Test x Condition interaction, $F(3,75) = 3.24$, $p < .05$, $MS_e = .53$, $\eta^2 = .08$. As shown in Fig. 2, social learning caused participants in the ambiguous and

untrustworthy conditions to judge faces as significantly less trustworthy while judging faces in the positive and mixed conditions as slightly more trustworthy. We found a trend toward a three-way interaction of Group, Test and Condition, $F(3,75) = 2.42$, $p = .07$, $MS_e = .53$, $\eta^2 = .06$, because the effects of training on condition were largely restricted to participants low in BPD features. An ANOVA of RTs yielded only one statistically significant effect: a main effect of Test, $F(1,25) = 9.47$, $p < .01$, $MS_e = .03$, $\eta^2 = .74$. Participants were significantly faster at rating face stimuli after (1036 ms) compared to before training (1112 ms). We found a trend toward a Test x Condition interaction, $F(3,75) = 2.46$, $p = .07$, $MS_e = .53$, $\eta^2 = .08$, because the speedup from training was relatively stronger in the negative condition and relatively weaker in the ambiguous condition.

ERP analyses. A mixed model ANOVA was performed separately on the P1, P2, N2 and slow wave ERP components, with Group (2 levels: low BPD, high BPD) as the between-subjects factor and Test (2 levels: pretest, posttest) and Condition (4 levels: ambiguous, negative, positive, mixed) as within-subjects factors. An effect of training was found in the P1 ERP component, $F(1,25) = 5.07$, $p < .05$, $MS_e = 261.73$, $\eta^2 = .55$, with the magnitude of P1 significantly larger at posttest (1.98 μV) than at pretest (.19 μV). However, there was no effect of Group, $F(1,25) = .01$, *ns*, $MS_e = 246.91$, $\eta^2 = .001$, and no interaction between Group and Test, $F(1,25) = .01$, *ns*, $MS_e = 261.73$, $\eta^2 = .001$, indicating that the sensory effects of training were equivalent between groups.

A trend-level effect of training was also found in the P2 ERP component, $F(1,25) = 3.90$, $p = .06$, $MS_e = 30.21$, $\eta^2 = .28$, with P2 amplitude at posttest (2.68 μV) diminished relative to pretest (3.69 μV). There was no main effect of Group, $F(1,25) = .18$, *ns*, $MS_e = 284.51$, $\eta^2 = .12$. However, there was a trend-level interaction between Group and Test, $F(1,25) = 3.790$, $p = .06$, $MS_e = 30.21$, $\eta^2 = .27$: The drop in P2 after training was relegated to the high BPD participants. Similar trends were found in the N2 ERP component. There was a significant main effect of Test, $F(1,25) = 5.06$, $p < .05$, $MS_e = 68.47$, $\eta^2 = .28$, with N2 amplitude growing more negative from pretest (-.23 μV) to posttest (-.52 μV). There was no effect of Group, $F(1,25) = 1.12$, *ns*, $MS_e = 395.12$, $\eta^2 = .35$, but Group and Test trended toward interaction, $F(1,25) = 2.51$, $p = .13$, $MS_e = 68.47$, $\eta^2 = .14$, because the effects of training were seen only in the high BPD group.

A different pattern was observed in analysis of the slow wave. Here, we found no effect of Test, $F(1,25) = .45$, *ns*, $MS_e = 107.05$, $\eta^2 = .02$, or Group, $F(1,25) = 2.71$, *ns*, $MS_e = 376.18$, $\eta^2 = .38$, and no interaction between the two variables, $F(1,25) = .24$, *ns*, $MS_e = 107.05$, $\eta^2 = .01$. However, there was a significant three-way interaction with Test, Group, and Condition, $F(3,75) = 3.40$, $p < .05$, $MS_e = 28.02$, $\eta^2 = .11$. As one can see in Fig. 3, after training the slow wave decreased in magnitude for low BPD participants in the negative condition, but decreased in magnitude for high BPD participants in the positive and mixed conditions.

Discussion

The results of the current study replicate the untrustworthiness bias found previously in patients with BPD (Fertuck, Grinband, & Stanley, 2013; Fertuck et al., 2019), in a group of non-clinical undergraduates who endorsed features of the disorder compared to those with few features. Participants high in BPD features rated neutral faces as significantly more untrustworthy than participants low in BPD symptoms. Moreover, we were able to demonstrate for the first time that social learning can be used to significantly modify trust appraisals in participants high and low in BPD features. We found that providing participants with untrustworthiness or ambiguous trait descriptors enhanced their ratings of untrustworthiness to neutral faces, whereas providing them with positive or mixed descriptors yielded higher ratings of trustworthiness.

The neural effects of training were observed in both early and late ERP components. The P1 component, occurring approximately 50 ms after stimulus onset, was enhanced in magnitude after training, equally for both groups of participants. The slow wave ERP component, a measure of sustained attention occurring 700–900 ms after stimulus onset (Bidet-Caulet et al., 2010; Näätänen et al., 1978, 1982), was affected differentially by condition and group. L-BPD participants revealed greater negative slow wave activity after training (i.e., more sustained attention) to stimuli with negative trait descriptors, whereas H-BPD participants showed greater focus after training to faces accompanying positive and mixed descriptors. The results suggest that social learning affects participants with BPD features at both the neural and behavioral levels.

Model of BPD trustworthiness appraisal processes

In duplicating the untrustworthiness bias in participants high in BPD features, the results of the current study suggest that investigations of non-clinical samples may serve as a useful model of appraisal processes in BPD, at least as a first approximation. Our results indicate that in endorsing BPD features, healthy undergraduates share a bias with BPD patients for appraising neutral faces as untrustrworthy, compared with those who endorse relatively few BPD features. One conceptualization places BPD along a continuum of symptom severity, with both BPD patients and H-BPD sharing key symptoms related to trust appraisal, such as rejection sensitivity. On this view, individuals with H-BPD, like BPD, feel excluded in social settings and react with fear or anger, creating a deep-seated sense of mistrust in others, a tendency revealed in trust appraisals of neutral faces. Research on healthy populations with BPD features could now be extended to other trust paradigms, including neuroeconomic games, to further explore similarities and differences of impairments in trust appraisal between H-BPD and BPD.

Behavioral effects of social learning

We found that social learning was effective in both increasing and decreasing trust appraisals. The behavioral effects were numerically larger for L-BPD than H-BPD, in keeping with the view that BPD features are resistant to change. Nevertheless, it is noteworthy that overall effects of training were found in both ratings and RTs, indicating that short-term training is effective in changing trust appraisals even in individuals with high levels of BPD symptoms. Although the results of this preliminary study are

suggestive of social learning as a new approach to trust bias training, follow-up investigations using the approach with BPD patients are needed to explore feasibility in clinical settings.

Neuroplasticity from social learning

Unlike previous studies of BPD patients versus healthy controls (e.g., Marissen, Meuleman, & Franken, 2010; Meares, Melkonian, Gordon, & Williams, 2005; Meares, Schore, & Melkonian, 2011; Ruchsow et al., 2008), we found no main effects of Group in any ERP measures time locked to behavioral responses: P1, P2, N2 and slow wave. Thus, despite identifying a mistrust bias in ratings, we were unable to pinpoint a corresponding neural signature of the bias. Although our decision to measure these four ERP components was based on previous research, no extant study has examined ERP correlates of the untrustworthiness bias. Thus it is conceivable that a neural signature of the bias resides in a different ERP component.

Nevertheless, we uncovered ample evidence of social learning on neural processes, beginning with the P1 ERP component. Here, we found that training enhanced P1 amplitude for both groups. We also found trends after training in both the P2 (smaller after training) and N2 (larger after training) ERP components, both of which were numerically greater in the H-BPD group. These results indicate that, at the neural level, H-BPD is at least as amenable to change from learning as L-BPD, despite showing weaker behavioral change. Previous research suggests that behavioral and electrophysiological measures follow different time courses in response to cognitive training (Atienza et al., 2002; Atienza et al., 2005; Tremblay, et al., 1998), perhaps reflecting different learning mechanisms (Ahissar and Hochstein, 2004). Future research incorporating posttest measures at distinct time intervals may help to better characterize the time course of neuroplasticity in BPD.

Group-specific neural effects were relegated in the current study to the late slow wave component, occurring 700–900 ms after the onset of the face stimuli. Theoretically, the slow wave component measured here is linked to *processing negativity* (PN), an ERP component that accompanies attended stimuli (Bidet-Caulet et al., 2010; Näätänen et al., 1978, 1982). Greater negativity in PN is associated with greater sustained attention to task-relevant stimuli (Näätänen, 1992). Hence, the three-way interaction of Group, Test and Condition uncovered in the current study may indicate group differences in the effects of training on trait descriptors. Specifically, L-BPD participants revealed greater slow-wave negativity after training to negative trait descriptors, whereas H-BPD participants showed greater slow-wave negativity to positive and mixed descriptors. The effects of training on ratings to these descriptors were opposite, with negative descriptors evincing higher ratings of untrustworthiness and positive and mixed descriptors evincing higher ratings of trustworthiness. Perhaps then the slow-wave effects of training correspond to group differences in attentional effort required for behavioral change, with H-BPD necessitating greater effort to overcome an inherent mistrust bias.

Conclusions

The current study is the first to show both neural and behavioral effects of social learning on the untrustworthiness bias in individuals endorsing borderline features. Trait descriptors were effective in changing trust judgments in both positive and negative directions. Neuroplasticity of training began as early as 50 ms after stimulus onset but only revealed group differences 700 ms after stimulus onset. Our results point to the possibility of developing effective therapeutic interventions for the mistrust bias. Future research should examine the social learning paradigm in a clinical population of those with BPD compared to healthy and psychiatric controls.

Declarations

Ethical Approval and Consent to participate: The study was approved by the IRB of the City University of New York, and all participants consented.

Consent for publication: Not applicable.

Availability of supporting data: Supporting data is available for appropriate use.

Competing interests: The authors have no competing interests

Funding: This study was supported in part by a grants from City Seed of the City University of New York.

Authors' contributions: E.F., S.F., and R.M. have contributed to writing, stimulus development, data collection, data analysis, and interpretation within this submission. E.F. and R.M. contributed to the development of the supporting grant and intellectual conceptualization of the project.

Acknowledgments: Rafal Skiba, Raquel Bibi, Rasheda Browne, Kwesi Sullivan, Stephanie Fisher, Estephanie Bravo, Naomi Dambreville, Jay Edelman, Esen Karan, Neelam Prashad, Aashna Shah, Ana Kodra, and Grace John contributed to participant recruitment, data collection, administration, and stimulus development.

References

1. Adolphs R, Tranel D, Damasio AR. The human amygdala in social judgment. Nat Rev Neurosci. 1998;393:470–4.
2. Ahissar M, Hochstein S. The reverse hierarchy theory of visual perceptual learning. Trends in Cognitive Science. 2004;8:457–64.
3. Amaral DG. The primate amygdala and the neurobiology of social behavior: implications for understanding social anxiety. Biol Psychiatry. 2002;51(1):11–7.
4. Atienza M, Cantero JL, Dominguez-Marin E. The time course of neural changes underlying auditory perceptual learning. Learning Memory. 2002;9:138–50.
5. Atienza M, Cantero JL, Quiroga RQ. Precise timing accounts for posttraining sleep-dependent enhancements of the auditory mismatch negativity. Neuroimage. 2005;26:628–34.

6. Barnow S, Stopsack M, Grabe HJ, Meinke C, Spitzer C, Kronmuller K, Sieswerda S. Interpersonal evaluation bias in borderline personality disorder. Behav Res Ther. 2009;47(5):359–65. doi:10.1016/j.brat.2009.02.003.
7. Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron. 2008;58(4):639–50. doi:10.1016/j.neuron.2008.04.009.
8. Beck AT, Butler AC, Brown GK, Dahlsgaard KK, Newman CF, Beck JS. Dysfunctional beliefs discriminate personality disorders. Behav Res Ther. 2001;39:1213–25.
9. Berchio C, Piguet C, Gentsch K, Küng A-L, Rihs TA, Hasler R,.. . Perroud N. Face and gaze perception in borderline personality disorder: An electrical neuroimaging study. Psychiatry Research: Neuroimaging. 2017;269(Supplement C):62–72. doi:https://doi.org/10.1016/j.pscychresns.2017.08.011.
10. Bidet-Caulet A, Mikyska C, Knight RT. Load effects in auditory selective attention: Evidence for distinct facilitation and inhibition mechanisms. NeuroImage. 2010;50:277–84.
11. Diamond A, Lee K. Interventions shown to aid executive function development in children 4 to 12 years old. Science. 2011;333(6045):959–64.
12. Engell AD, Haxby JV, Todorov A. Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala. J Cogn Neurosci. 2007;19(9):1508–19. doi:10.1162/jocn.2007.19.9.1508.
13. Fertuck EA, Fischer S, Beeney J. Social Cognition and Borderline Personality Disorder: Splitting and Trust Impairment Findings. Psychiatr Clin North Am. 2019;41(4):613–32. doi:10.1016/j.psc.2018.07.003.
14. Fertuck EA, Grinband J, Stanley B. Facial trust appraisal negatively biased in borderline personality disorder. Psychiatry Res. 2013;27:195–202.
15. Fertuck EA, Grinband J, Mann JJ, Hirsch J, Ochsner K, Pilkonis P,.. . Stanley B. Trustworthiness appraisal deficits in borderline personality disorder are associated with prefrontal cortex, not amygdala, impairment. Neuroimage Clin. 2019;21:101616. doi:10.1016/j.nicl.2018.101616.
16. Garner KG, Matthews N, Remington RW, Dux PE. Transferability of training benefits differs across neural events: Evidence from ERPs. J Cogn Neurosci. 2015;27:2079–94.
17. Gutz L, Renneberg B, Roepke S, Niedeggen M. Neural processing of social participation in borderline personality disorder and social anxiety disorder. J Abnorm Psychol. 2015;124(2):421–31. doi:10.1037/a0038614.
18. Koenigsberg BT, Denny, Jin F, Liu X, Guerreri S, Mayson SJ,.. . Larry J, Siever. The Neural Correlates of Anomalous Habituation to Negative Emotional Pictures in Borderline and Avoidant Personality Disorder Patients. Am J Psychiatry. 2014;171(1):82–90. doi:10.1176/appi.ajp.2013.13070852.
19. Jaeggi SM, Buschkuehl M, Shah P, Jonides J. The role of individual differences in cognitive training and transfer. Memory Cognition. 2014;42(3):464–80.

20. Meares R, Melkonian D, Gordon E, Williams L. Distinct pattern of P3a event-related potential in borderline personality disorder. Neuroreport. 2005;16(3):289–93.
21. Meares R, Schore A, Melkonian D. Is Borderline Personality a Particularly Right Hemispheric Disorder? A Study of P3a Using Single Trial Analysis. Australian New Zealand Journal of Psychiatry. 2011;45(2):131–9. doi:10.3109/00048674.2010.497476.
22. Melara RD, Singh S, Hien DA. (2018). Neural and behavioral correlates of attentional inhibition training and perceptual discrimination Training in a Visual Flanker Task. *Frontiers in Human Neuroscience*, *12*.
23. Millner AJ, Jaroszewski AC, Chamarthi H, Pizzagalli DA. Behavioral and electrophysiological correlates of training-induced cognitive control improvements. Neuroimage. 2012;63(2):742–53.
24. Näätänen R. Processing negativity: an evoked-potential reflection of selective attention. Psychology Bulletin. 1982;92:605–40.
25. Näätänen R. Attention and brain function. Hillsdale: Erlbaum; 1992.
26. Näätänen R, Gaillard AWK, Mantysalo S. Early selective attention effect on evoked potential reinterpreted. Acta Physiol (Oxf). 1978;42:313–29.
27. Rueda MR, Checa P, Cómbita LM. Enhanced efficiency of the executive attention network after training in preschool children: immediate changes and effects after two months. Dev Cogn Neurosci. 2012;2:192–204.
28. Said CP, Baron SG, Todorov A. Nonlinear Amygdala Response to Face Trustworthiness: Contributions of High and Low Spatial Frequency Information. J Cogn Neurosci. 2008;21(3):519–28. doi:10.1162/jocn.2009.21041.
29. Schmidt NB, Richey JA, Buckner JD, Timpano KR. Attention training for generalized social anxiety disorder. J Abnorm Psychol. 2009;118:5–14.
30. Tang YY, Posner MI. Attention training and attention state training. Trends in Cognitive Sciences. 2009;13(5):222–7.
31. Thorell LB, Lindqvist S, Bergman Nutley S, Bohlin G, Klingberg T. Training and transfer effects of executive functions in preschool children. Developmental Science. 2009;12(1):106–13.
32. Tremblay KL, Kraus N, McGee T. The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. NeuroReport. 1998;9:3557–60.
33. Winston JS, Strange BA, O'Doherty J, Dolan RJ. Automatic and intentional brain responses during evaluation of trustworthiness of faces. Nat Neurosci. 2002;5(3):277–83.

Figures

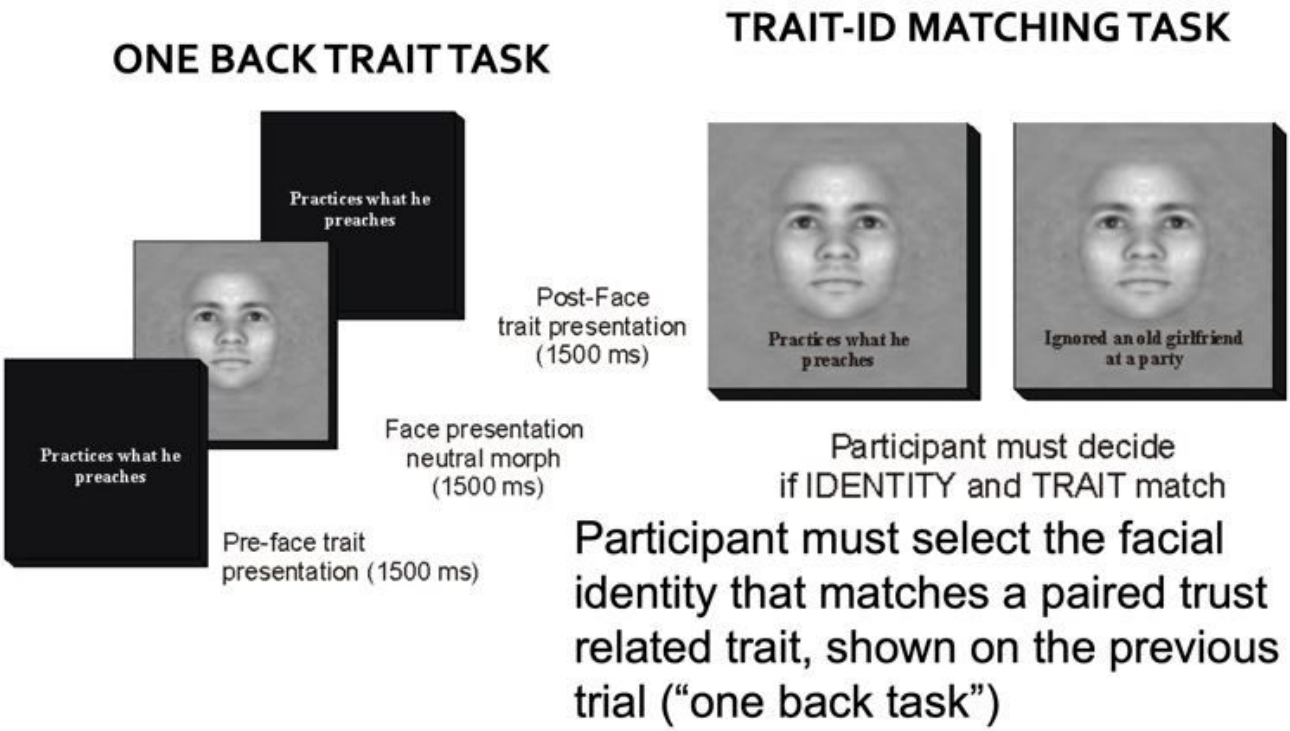
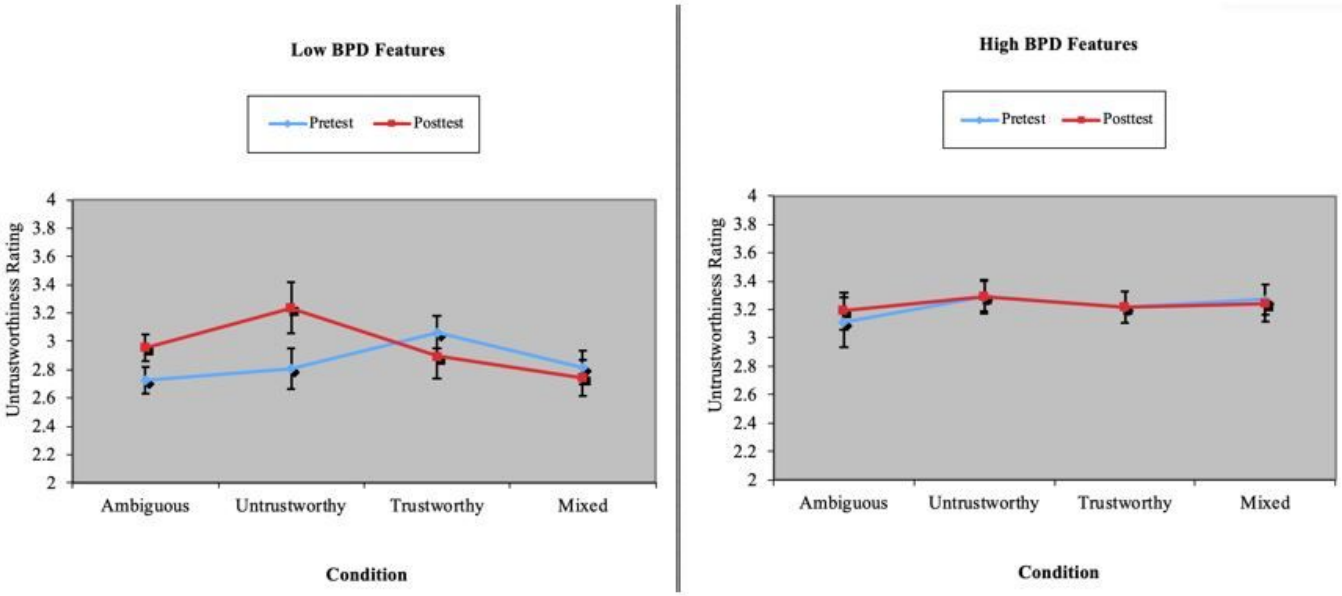


Figure 1

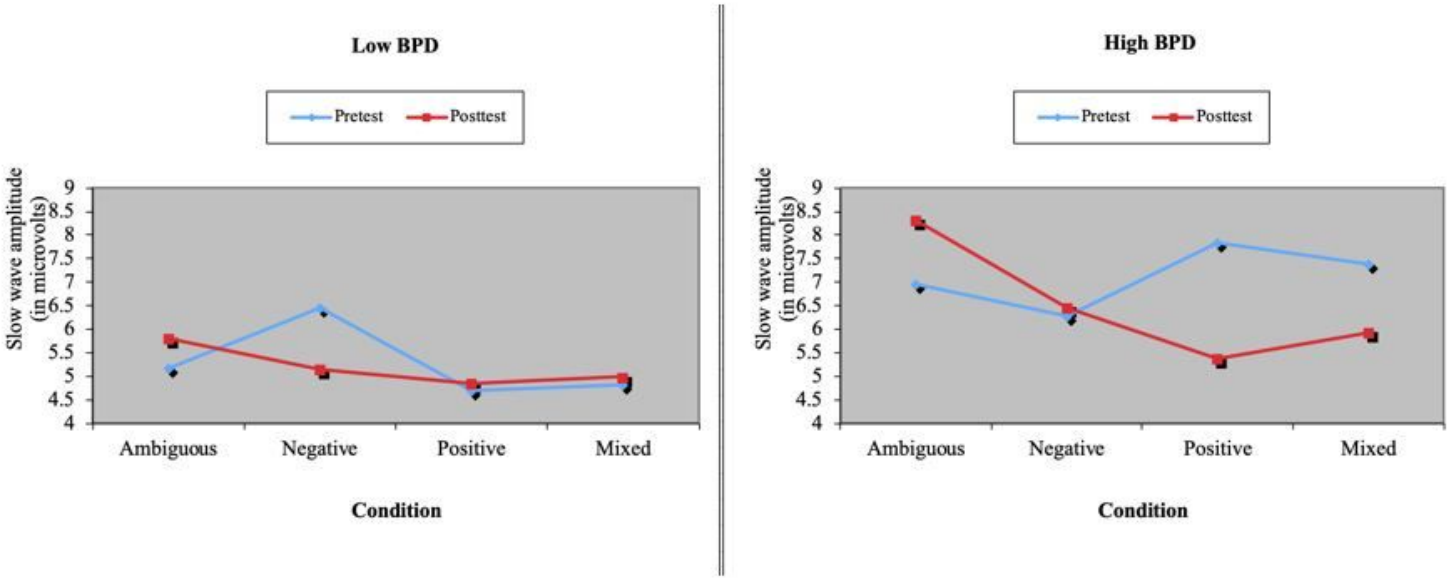
Trustworthiness Trait Learning Task.



Note. Left panel (Low BPD) and right panel (High BPD) on mean trustworthiness ratings (y axis) and learning condition (x axis).

Figure 2

Trustworthiness ratings Pre and Post Social Learning in High and Low BPD Features Groups by Condition.



Note. Left panel (Low BPD) and right panel (High BPD) on mean Slow Wave ERP (y axis) and learning condition (x axis).

Figure 3

Evoked Response Potential (ERP) Slow Wave ratings Pre and Post Social Learning in High and Low BPD Features Groups by Condition.