

Trust beliefs, biases, and behaviors in borderline personality disorder: Empirical findings and relevance to epistemic trust

Sara Rose Masland, Ph.D.*

Sophie Elizabeth Schnell

Tanya Vikram Shah

Pomona College Department of Psychological Science

*Corresponding author: Pomona College Department of Psychological Science, 647 North College Way, Claremont CA 91711; sara.masland@pomona.edu

Masland, S.R., Schnell, S.E., & Shah, T.V. (in press). Trust beliefs, biases, and behaviors in borderline personality disorder: Empirical findings and relevance to epistemic trust. *Current Behavioral Neuroscience Reports*.

Abstract

Purpose of Review: This review summarizes empirical research on trust in BPD, including three primary areas: the prevalence of paranoia, trustworthiness appraisals, and trust-related behaviors in economic exchange paradigms. Connections to the largely theoretical study of epistemic trust in BPD are highlighted.

Recent Findings: In trust appraisal paradigms, people with BPD have a bias to rate others as untrustworthy. In behavioral exchange games, they report lower trust in partners and are more likely to rupture cooperation. Recent research has suggested potential explanations for these findings, including differences in affective processing, aberrant social norms and expectations, and difficulty attending to and incorporating social cues.

Summary: People with BPD commonly experience paranoia, generally regard others as untrustworthy, and act accordingly. Future research is needed to understand the mechanisms of altered trust processing and to integrate empirical research with recent theoretical research on epistemic trust.

Keywords: borderline personality disorder; trust; cooperation; epistemic trust

Introduction

In every interaction—from sharing an elevator with a stranger to discussing relationship difficulties with a spouse—trust enables us to proceed in an inherently uncertain world. To trust is to make oneself interpersonally vulnerable, with the expectation and/or hope that others will act with positive intentions [1]. Trust is central to interpersonal interactions across contexts, and particularly relevant to close relationship functioning [2]. Trust difficulties predict lower relationship satisfaction [3], greater variability in perceptions of romantic partners [4], and lower perceived closeness to partners [5].

Difficulties with trust are particularly salient in people with borderline personality disorder (BPD). Although typical descriptions of BPD do not feature trust impairments, they are implied in the disorder's diagnostic criteria. BPD includes temporary, stress-related paranoia, as well as significant fears of interpersonal abandonment and unstable representations of relationship partners [6]. In the Alternative Model of Personality Disorders [6], BPD is defined in part by high trait antagonism, akin to low agreeableness, a trait domain that includes trust [7].

Trust difficulties in people with BPD are of longstanding clinical interest [8–16]. Research on trust in BPD has focused predominantly on three main areas: 1) paranoia and suspiciousness; 2) appraisals of the trustworthiness of others; and 3) trust-related behaviors in economic exchange paradigms and the influence of oxytocin on these behaviors. A consistent message emerges from these studies: people with BPD commonly consider others untrustworthy and act accordingly. In this review, we consider recent empirical findings related to trust processing in BPD across these three areas of study—beliefs, biases, and behavior. Additionally, we consider an emerging body of largely theoretical research on epistemic trust, and suggest how empirical findings support its relevance to BPD.

Paranoia and Mistrust Beliefs

The term “borderline” was originally coined to connote that the disorder exists in a liminal space between neurosis and psychosis [17]. Although many researchers and clinicians currently use the term without meaning to imply this theoretical stance on borderline pathology, the relevance of psychosis to BPD remains. Quasi-psychotic thought has been called “virtually pathognomonic” in BPD because of its 100% prevalence in one study [18]. The experience of quasi-psychotic thought in BPD often centers on nondelusional paranoia (NDP), which includes interpersonal distrust and a belief in the malevolent intentions of others [18–20]. NDP strongly predicts poorer personal and social functioning in BPD [19] and mediates the association of early maltreatment and suicide potential [21]. Although the experience of psychosis and paranoia is often regarded as nondelusional, it warrants mention that a smaller proportion of those with BPD do experience paranoid delusions [19].

In one longitudinal study, Zanarini et al. [20] found that, at baseline, 86.6% of people with BPD endorsed NDP, compared to 43.1% of people with other personality disorders (PDs). Although NDP decreased significantly for both groups over follow-up assessments, the BPD group endorsed NDP at more than twice the rate (43.3%) of those with other PDs (20.7%) 16 years later. More recent work has found that although paranoia is not uncommon in other personality disorders, it is particularly associated with risk for BPD [22]. Moreover, paranoia is more prevalent and severe in BPD compared to dysthymia [23] and even schizophrenia [19].

Given that the experience of paranoia is more common in BPD and purportedly “transient” and “stress-related” [6], one might expect people with BPD to have a greater degree of paranoid reactivity in the context of stress, relative to other groups. However, there is limited and mixed evidence for this difference. In an experience sampling study, Glaser et al. [24] found

that people with BPD showed the greatest psychotic reactivity to daily stress compared to people with Cluster C PDs, people with psychotic disorders, and healthy controls. However, although specifically *paranoid* reactivity was greater for patient groups relative to healthy controls, it did not differ between patient groups.

Perhaps unsurprisingly given the prevalence of NDP in BPD, schemas that include mistrust of others figure prominently in the disorder. These include distrust [25], mistrust/abuse [26,27], and disconnection/rejection [28]. The prevalence of mistrust/abuse beliefs may differentiate BPD from other personality disorders (e.g., [26,29]), and it has been suggested that the predominance of mistrustfulness/anger beliefs may define a subtype of BPD [30].

The specific manner in which paranoia and mistrust beliefs in BPD manifest in social judgments and interactions is not clear from studies of prevalence and severity alone. To better understand and isolate trust processes in people with BPD, researchers have turned to the lab, where biases in basic trust appraisals consistent with paranoia and mistrust are clear.

Trust Appraisals

In general, when provided with limited information—photographs or short videos—about others, people with BPD describe them more negatively compared to healthy controls and people with Cluster C PDs [31], and as more aggressive compared to healthy controls and people with unipolar depression [32]. For trustworthiness judgments specifically, it is now a well-replicated finding that people with BPD make more negative appraisals than other groups (e.g., [33–37]). Trust appraisal findings are summarized in Table 1.

In an early study on trustworthiness appraisals in BPD, people with BPD rated faces as less trustworthy and less approachable compared to healthy controls [36]. Using similar stimuli, Fertuck et al. [33] showed participants a series of faces that varied on either a fear or

trustworthiness dimension, and asked them to rate the fearfulness or trustworthiness of each face. Compared to controls, people with BPD had a slower response time for faces with ambiguous trustworthiness and generally rated male faces as less trustworthy but no more or less fearful. Their responses were consistent with a negative trustworthiness response bias, not with greater sensitivity to cues of trustworthiness or ability to discriminate trustworthy from untrustworthy faces. Negative response bias operated independently of fear appraisals, indicating that trustworthiness bias does not result from difficulty with fear recognition.

Some research has examined how negative emotional context influences trustworthiness appraisals. Although an induced negative emotional state does not seem to alter how people with BPD appraise the general malevolence of others [38], negative emotional primes may unduly influence trustworthiness appraisals for people with BPD. In a non-clinical sample, Masland and Hooley [34] asked participants to rate the trustworthiness of faces following neutral, negative, or positive affective primes. Regardless of prime condition, a BPD features group (people with three or more BPD criteria) rated faces as less trustworthy relative to a control group of people with two or fewer BPD criteria. This finding is consistent with a negative response bias. The BPD features group was also more greatly influenced by negative primes than the control group. Negative affect may augment the trust appraisal bias observed in BPD, and negative affective processes may shed light on affective mechanisms through which BPD influences trust.

Another potential mechanism is rejection sensitivity: a readiness to expect, perceive, and overreact to signs of rejection [39]. People with BPD have significant rejection sensitivity on self-report (e.g., [40,41]) and behavioral (e.g., [42]) measures. In non-clinical samples, rejection sensitivity mediates the association of BPD features with negative trustworthiness appraisals [35,37]. More specifically, the emotional components of rejection sensitivity (i.e., anxiety and

anger related to rejection) mediate the association, while the cognitive components (i.e., expectations of rejection) do not [37]. This again suggests that emotional processes contribute to trustworthiness appraisal.

The assessment of basic trustworthiness appraisals is conducive to experimentation, yet its ecological validity is questionable. Notably, trustworthiness appraisals pair participants with strangers, failing to account for the effect of relationship intimacy. In a rare study matching participants with close others, Miano et al. [43] examined trust appraisals in two sets of unmarried heterosexual couples: in one set the women were diagnosed with BPD; in the other set neither partner had BPD. Each couple discussed personal fears, possible partner separation, and favorite films. Trustworthiness appraisals of partners made after the discussion of favorite films did not differ for women with BPD versus healthy controls. Women with BPD showed a significant reduction in trustworthiness (but not confidence) appraisals of their partners following the threatening discussions. This is further evidence that general negative emotional context negatively impacts how women with BPD consider trustworthiness.

Behavioral Exchange Paradigms

The “trust game” is the most widely used and adapted behavioral paradigm for examining trust and cooperation in BPD. Borrowed from behavioral economics, it tracks financial decision-making in a simulated interaction [44]. Numerous trust game studies have reported decreased trust and cooperation in people with BPD [10,44,45]. Findings from behavioral exchange paradigms are summarized in Table 1.

King-Casas et al. [44] were the first to study trust game behavior in people with BPD. They employed a multi-round economic exchange game in which an investor and a trustee each begin with a set sum of money. A single round proceeds as follows: the investor decides

how much money to send to the trustee, this amount is tripled and transferred to the trustee, the trustee elects what proportion of this updated sum (the initial sum plus the received investment) to return to the investor, and the chosen amount is transferred back to the investor. Cooperation from both parties is mutually beneficial. Investors demonstrate trust in their partner by sending large investments; trustees reciprocate trust by making a large repayment. Compared to healthy trustees, trustees with BPD were less trusting and cooperative in the exchange game [44]. They reported lower levels of trust in investors and, accordingly, made significantly smaller repayments. Although investors initially transferred the same amount of money to healthy trustees and trustees with BPD—reflecting equal trust in both groups—they transferred significantly less to trustees with BPD in later rounds—as the game progressed, investors lost trust in trustees with BPD due to their uncooperative behavior.

Later trust game studies distinguished active cooperation (fairness, non-exploitation) from reactive cooperation (forgiveness, non-retaliation). Thielmann et al. [46] found that individuals with higher levels of borderline features displayed a greater tendency to retaliate, suggesting that BPD impairs only *reactive* cooperation—hindering the ability to tolerate and forgive. Hepp et al. [47] found no differences in active cooperation between people with BPD and healthy controls. Thus, individuals with BPD do not transfer less money because they are more exploitative or unfair.

Other work has considered that heightened risk aversion, as opposed to diminished interpersonal trust, may explain decreased spending in people with BPD. Unoka et al. [45] conducted both a traditional trust game and a “risk game,” in which participants believed that a computer lottery program (instead of a human partner) randomly determined repayments. In the trust game, investors with BPD transferred significantly less money to trustees and predicted

worse outcomes. However, in the risk game, investors with BPD and controls transferred comparable amounts of money. Together, these findings demonstrate that decreased spending specifically reflects diminished interpersonal trust. The finding that BPD affects trusting behavior and not risk-aversion has been replicated [48,49].

Trust game data establish that people with BPD are less trusting and suggest explanations for why this is the case. One possible explanation, supported by multiple studies, is the possession of atypical social norms [44,45,50]. While playing an economic exchange game, people with BPD exhibited differential activity in the bilateral anterior insula, which is known to respond to social norm violations [44]. In healthy controls, bilateral anterior insula activity increased after receiving small investments, which violated the social norms of trust and cooperation. Large, norm-adhering investments provoked less insula activity. Unlike healthy trustees, those with BPD showed similar insula activity in response to both low and high investments. As they hold distorted, overly negative social expectations, trustees with BPD may anticipate partner betrayal and, therefore, may perceive low investments as normative.

Critically, evidence from trust games suggests that people with BPD predict partner betrayal regardless of actual partner behavior. Liebke et al. [50] found that following rejection, both healthy controls and people with BPD adjusted their expectations to predict future rejection. However, following acceptance, only the healthy controls adjusted their expectations to predict more positive feedback. People with BPD did not become more trusting after an experience of social acceptance. In fact, people with BPD who had been socially accepted were actually less trusting of cooperative trustees. Liebke et al. [50] speculated that the ability to learn from positive social experiences—and to adjust one's social expectations and behaviors accordingly—may be compromised in people with BPD.

The inability to use partner feedback to adjust behavior is disadvantageous when a partner is trusting: the failure to reciprocate ruptures trust in the partnership. This inability is also disadvantageous when a partner is distrustful, as it prevents trust from being repaired. In economic exchange games, participants have the opportunity to repair trust by “coaxing.” When they receive a low investment—a cue that they’ve ruptured trust with their investor—participants can choose to repay a high proportion to demonstrate their renewed trustworthiness and secure larger investments in subsequent rounds. Healthy trustees are more likely than trustees with BPD to coax following a low investment [44]. Trustees with BPD do not change their behaviors in order to regain investor trust and repair the alliance. In a re-analysis of the data set from King-Casas et al. [44], Hula et al. [48] identified difficulties adapting to investor irritation in a subgroup of people from both BPD and healthy control groups, whom they labeled “perilous.” A significantly larger proportion of these perilous trustees came from the BPD group. Perilous trustees were more likely to rupture trust with their investors and to fail to repair the alliance.

Traditional trust games are not the only context in which people with BPD experience difficulties with social learning. Using the Social Valuation Task (SVT), an economic-decision making game in which participants receive trustworthy and untrustworthy advice from an actor posing as another participant, Fineberg et al. [51] found that although people with BPD were more attentive to social (and nonsocial cues), they had slower social learning rates compared to healthy controls. In conditions where actor advice was particularly volatile, people with BPD adapted their learning more slowly and were less sensitive to increased volatility. Fineberg et al. [51] suggested that people with BPD may expect greater volatility (or interpersonal unreliability), and thus are not as surprised by or as sensitive to rapid changes in behavior or trust-related cues from others. In further support of this idea, people with BPD exhibited less

implicit distress upon learning that their partner was actually an actor in the post task debriefing. Confirmation of partner deception seemingly reduced distress in people with BPD.

People with BPD may be less able to adapt when they encounter volatility and irritability, because these cues confirm their expectations. Alternatively, impairments in mentalizing, the ability to understand the mental states of oneself and others [52] may explain apparent difficulties incorporating social cues in trust-related behaviors. Although some argue that trust games data indicates superior mentalizing in people with BPD [53], BPD is commonly understood as a disorder of impaired mentalizing [52]. On a neural level, disrupted insula signaling in people with BPD could hinder awareness that a relationship is in jeopardy and in need of repair [54].

Variance in levels of awareness and cooperation among people with BPD demonstrates that trusting behaviors across individuals with BPD are not uniform, and highlights the importance of individual differences. Although Hula et al. [48] found that “perilous” trustees were overrepresented in the BPD group, it is crucial to note that not all individuals with BPD qualified as “perilous.” In fact, with the exception of this subgroup, no significant behavioral differences were found between people with BPD and healthy controls. Hula et al. [48] proposed the development of a “perilous” clinical phenotype to distinguish this subgroup categorically from people with BPD who are more cooperative. From a dimensional standpoint, personality traits associated with BPD (i.e., high neuroticism and low agreeableness) also predict a failure to repair broken trust in exchange games [46,55]. Individual differences with respect to personality traits should be further considered.

The Effect of Oxytocin

Because the neuropeptide oxytocin (OXT) increases trusting behavior in healthy individuals [56–58], it seems to hold some promise for treating disorders in which interpersonal trust and cooperation is compromised, including BPD. However, across trust game studies, intranasal OXT administration actually reduces trust in people with BPD [59,60]. In a game requiring mutual cooperation (as opposed to defection) for maximum monetary outcome, people with BPD who received OXT trusted their partners significantly less than people with BPD who received a placebo [59]. Inpatients with BPD were also less trusting in a traditional economic exchange game when administered OXT [60].

OXT differentially affects healthy controls and people with BPD. However, even among people with BPD, the effect of OXT is moderated by individual differences, including variation in attachment anxiety and attachment avoidance. In one study, “fearful” individuals with both high attachment anxiety and avoidance were less cooperative when administered OXT, while “preoccupied” individuals with high anxious attachment and low attachment avoidance were more cooperative when administered OXT [59]. Additionally, childhood adversity may moderate OXT’s effects. In a trust game, inpatients with a history of early emotional neglect transferred significantly less money when administered oxytocin compared to placebo [60].

There are several possible explanations for the divergent effects of OXT on trust, both among people with BPD and between people with BPD and healthy controls. First, OXT may increase the salience of social cues (both positive and negative) rather than solely positive social emotions (e.g., trust; [56,59]). Therefore, for people with BPD, anxious attachment, and/or a history of childhood trauma, exogenous OXT may exacerbate negative social emotions, which, in turn, lower trust. Second, if OXT increases feelings of connectedness, it may inspire fears of

abandonment and remembrances of past interpersonal conflict for these same populations [59]. OXT increases both in-group connection and out-group aggression [56]. Given difficulties with splitting and identity instability in BPD, it is possible that the distinction between in-group and out-group may be blurred in moments of distress. Accordingly, OXT administration may increase aggression towards close others during interpersonal conflict. While this idea has yet to be investigated, it is consistent with the finding that, whereas healthy controls report greater trust in closer partners, people with BPD make no distinctions between peripheral and central members of their social networks in trust appraisals [61]. Lastly, people with BPD may have a dysregulated OXT system that reacts differently to exogenous OXT [59]. Most likely, some combination of these factors—in addition to others—underlies differential responses to OXT administration for people with BPD.

Recently, research has expanded beyond OXT to examine the effect of the drug acetaminophen on individuals with BPD. In people with high BPD features, acetaminophen reduces distrusting behavior in economic exchange games [62]. Future research should investigate treatment interventions—both pharmacological and psychotherapeutic—with the potential to promote trust in people with BPD.

Epistemic Trust

We are struck by the considerable unexplored connections between these empirical findings and recent, largely theoretical, work on epistemic trust. Epistemic trust is the belief that knowledge gained through social learning is authentic, relevant to oneself, and generalizable [63]. The development of epistemic trust enables the transmission of important social and cultural information. Difficulties with epistemic trust interfere with salutogenesis, the ability to derive positive benefit from social experiences [8,64]. Lack of salutogenesis creates generalized

issues with resilience to life stressors, which predispose an individual to psychopathology broadly and to BPD in particular [8,65].

The common experience of paranoia and the bias to rate others as less trustworthy are congruent with the general manifestation of epistemic mistrust, which includes the assumption of malevolent intentions of others [63]. Behavioral exchange studies are also consistent with epistemic mistrust and lack of salutogenesis in BPD. For example, the failure to update betrayal expectations after the experience of acceptance [50] demonstrates a failure to update existing belief systems in a manner that would be beneficial, as does the failure to invest in others or heed their advice, even when doing so would result in the best outcome.

Moreover, developmental research and theory suggests that epistemic trust requires openness to ostensive cues from interaction partners [66]. To experience salutogenesis, one must be receptive to the social cues that indicate an interaction partner is attempting to share information. Empirical findings on both trustworthiness appraisal and behavior in exchange games suggest impaired ability to use social cues for personal benefit. In multiple studies, people with BPD were seemingly insensitive to cues of irritability from others [44,48,51]. They may also use emotional cues to a lesser extent [53], or use inappropriate cues, including partner attractiveness [60]. It is possible that these findings are related to epistemic vigilance—the essential belief or expectation that others have misinformation [67]. These findings may also help explain why oxytocin, which increases the salience of social cues and should seemingly increase openness to social information, further entrenches mistrust for people with BPD. If social cues are threatening or reminiscent of early learning experiences that resulted in harm, they may signal that epistemic trust is maladaptive.

Evidence that adverse childhood experiences influence trust processing for people with BPD is also consistent with epistemic trust theory. In both trust appraisal studies and behavioral exchange games, childhood adversity predicts trust processing alterations [33,60]. In qualitative studies, people with BPD frequently link attachment disruption or abuse to their difficulties trusting others [11,13]. Attachment disruptions, including childhood neglect or abuse, are theorized to cause epistemic freezing or petrification, in which an individual is hypervigilant to interpersonal hazards and their capacity for trust is shut down [68] in a manner that may be adaptive when the social environment is dangerous [63,69].

Recently, MBT thought leaders have argued that the therapeutic relationship in MBT provides a new attachment experience in which patients can learn to build epistemic trust and openness, which may generalize to salutogenesis in social experiences outside of therapy [8,63]. Although the empirical study of epistemic trust is limited by the lack of measurement tools [15], initial evidence suggests that enhancing epistemic trust may improve outcomes for people with BPD [70]. Better operationalizations of epistemic trust in adolescents and adults will allow for fruitful integration of the types of empirical research reviewed here and treatment research.

Conclusion

Recent empirical research richly illustrates what has long been known clinically: people with BPD have difficulties trusting and cooperating with others. They commonly struggle with paranoia and/or mistrust beliefs, judge others as untrustworthy, and act accordingly—to their own detriment. However, the interplay of paranoia, beliefs, trust appraisals, trusting behaviors, and epistemic trust remains understudied. So long as these areas of research operate independently of one another, a comprehensive account of trust processing difficulties will remain obscured.

While this complex interplay may be best isolated and controlled in the laboratory, it is essential that trust research becomes more ecologically valid. Although trust games allow players to participate in simulated interactions, these interactions are far from realistic social encounters. Players interact solely with strangers, and the primary consequence of betrayed trust is economic, not social or personal. We urge the incorporation of real relationship partners into laboratory research. Outside of the lab, experience-sampling and daily-diary methods are well-suited to capture how real-world contexts influence trust processing. Research on trust in BPD also suffers from many of the limitations that pervade the broader BPD literature, including gender disparity, significant comorbidity, and medication differences between groups. These are detailed in Table 1.

Methodological limitations and variance notwithstanding, recent empirical efforts have done much to characterize how trust processing differences manifest in people with BPD. Future research should further refine that characterization, and consider the causes--both proximal and distal—and mechanisms underlying differences in trust processing. Additionally, the integration of work on trust beliefs, biases, and behavior is essential to improve not only our basic understanding of trust processing differences, but also patient care and outcomes. One particularly promising avenue for advancing patient care is the integration of these empirical efforts with largely theoretical work on epistemic trust, which may aid the development of trust-enhancing treatments for people with BPD.

References

1. Rousseau DM, Sitkin SB, Burt RS, Camerer C. Introduction to special topic forum: Not so different after all: A cross-discipline view of trust. *Acad Manage Rev* 1998;23:393-404.
2. Simpson JA. Psychological foundations of trust. *Curr Dir Psychol Sci* 2007;16:264-8. Doi:10.1111/j.1467-8721.2007.00517.x.
3. Fitzpatrick J, Lafontaine M-F. Attachment, trust, and satisfaction in relationships: Investigating actor, partner, and mediating effects. *Pers Relatsh* 2017;24:640-62. Doi:10.1111/pere.12203.
4. Campbell L, Simpson JA, Boldry JG, Rubin H. Trust, variability in relationship evaluations, and relationship processes. *J Pers Soc Psychol* 2010;99:14-31. Doi:10.1037/a0019714.
5. Kim JS, Weisberg YJ, Simpson JA, Oriña MM, Farrell AK, Johnson WF. Ruining it for both of us: The disruptive role of low-trust partners on conflict resolution in romantic relationships. *Soc Cogn* 2015;33:520-42. Doi:10.1521/soco.2015.33.5.520.
6. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington DC: 2013.
7. Trull TJ, Widiger TA. Dimensional models of personality: the five-factor model and the DSM-5. *Dialogues Clin Neurosci* 2013;15:135-46.
8. Bateman A, Campbell C, Luyten P, Fonagy P. A mentalization-based approach to common factors in the treatment of borderline personality disorder. *Curr Opin Psychol* 2018;21:44-9. Doi:10.1016/j.copsyc.2017.09.005.
9. Bo S, Sharp C, Fonagy P, Kongerslev M. Hypermentalizing, attachment, and epistemic trust in adolescent BPD: Clinical illustrations. *Personal Disord* 2017;8:172-82. Doi:10.1037/per0000161.
10. Botsford J, Schulze L, Bohländer J, Renneberg B. Interpersonal trust: Development and validation of a self-report inventory and clinical application in patients with borderline personality disorder. *J Pers Disord* 2019;1-22. Doi:10.1521/pedi_2019_33_462.
11. Fallon P. Travelling through the system: The lived experience of people with borderline personality disorder in contact with psychiatric services. *J Psychiatr Ment Health Nurs* 2003;10:393-400. Doi:10.1046/j.1365-2850.2003.00617.x.
12. Hirsh JB, Quilty LC, Bagby RM, McMain SF. The relationship between agreeableness and the development of the working alliance in patients with borderline personality disorder. *J Pers Disord* 2012;26:616-27. Doi:10.1521/pedi.2012.26.4.616.
13. Katsakou C, Pistrang N, Barnicot K, White H, Priebe S. Processes of recovery through routine or specialist treatment for borderline personality disorder (BPD): A qualitative study. *J Ment Health* 2019;28:604-12. Doi:10.1080/09638237.2017.1340631.

14. Langley GC, Klopfer H. Trust as a foundation for the therapeutic intervention for patients with borderline personality disorder. *J Psychiatr Ment Health Nurs* 2005;12:23-32. Doi:10.1111/j.1365-2850.2004.00774.x.
15. Orme W, Bowersox L, Vanwoerden S, Fonagy P, Sharp C. The relation between epistemic trust and borderline pathology in an adolescent inpatient sample. *Borderline Personal Disord Emot Dysregul* 2019;6:13. Doi:10.1186/s40479-019-0110-7.
16. Romeu-Labayen M, Rigol Cuadra MA, Galbany-Estragués P, Blanco Corbal S, Giralt Palou RM, Tort-Nasarre G. Borderline personality disorder in a community setting: Service users' experiences of the therapeutic relationship with mental health nurses. *Int J Ment Health Nurs* 2020. Doi:10.1111/inm.12720.
17. Stern A. Psychoanalytic Investigation of and Therapy in the Border Line Group of Neuroses. *Psychoanal Q* 1938;7:467-89.
18. Zanarini MC, Gunderson JG, Frankenburg FR, Chauncey DL. Discriminating borderline personality disorder from other axis II disorders. *Am J Psychiatry* 1990;147:161-7. Doi:10.1176/ajp.147.2.161.
19. Oliva F, Dalmotto M, Pirfo E, Furlan PM, Picci RL. A comparison of thought and perception disorders in borderline personality disorder and schizophrenia: Psychotic experiences as a reaction to impaired social functioning. *BMC Psychiatry* 2014;14. Doi:10.1186/s12888-014-0239-2.
20. Zanarini MC, Frankenburg FR, Wedig MM, Fitzmaurice GM. Cognitive experiences reported by patients with borderline personality disorder and axis II comparison subjects: A 16-year prospective follow-up study. *Am J Psychiatry* 2013;170:671-9. Doi:10.1176/appi.ajp.2013.13010055.
21. Allen B, Cramer RJ, Harris PB, Rufino KA. Borderline personality symptomatology as a mediator of the link between child maltreatment and adult suicide potential. *Arch Suicide Res* 2013;17:41-51. Doi:10.1080/13811118.2013.748413.
22. Muñoz-Negro JE, Prudent C, Gutiérrez B, Cervilla JA. Paranoia and risk of personality disorder in the general population. *Personal Ment Health* 2019;13:107-16. Doi:10.1002/pmh.1443.
23. Snyder S, Pitts WM. Characterizing paranoia in the DSM-III borderline personality disorder. *Acta Psychiatr Scand* 1986;73:500-5. Doi:10.1111/j.1600-0447.1986.tb02716.x.
24. Glaser J-P, Van Os J, Thewissen V, Myin-Germeys I. Psychotic reactivity in borderline personality disorder. *Acta Psychiatr Scand* 2010;121:125-34. Doi:10.1111/j.1600-0447.2009.01427.x.
25. Bhar SS, Brown GK, Beck AT. Dysfunctional beliefs and psychopathology in borderline personality disorder. *J Pers Disord* 2008;22:165-77. Doi:10.1521/pedi.2008.22.2.165.

26. Bach B, Farrell JM. Schemas and modes in borderline personality disorder: The mistrustful, shameful, angry, impulsive, and unhappy child. *Psychiatry Res* 2018;259:323-9. Doi:10.1016/j.psychres.2017.10.039.
 27. Bach B, Lobbestael J. Elucidating DSM-5 and ICD-11 diagnostic features of borderline personality disorder using schemas and modes. *Psychopathology* 2018;51:400-7. Doi:10.1159/000495845.
 28. Frías Á, Navarro S, Palma C, Farriols N, Aliaga F, Salvador A, et al. Early maladaptive schemas associated with dimensional and categorical psychopathology in patients with borderline personality disorder. *Clin Psychol Psychother* 2018;25:30-41. Doi:10.1002/cpp.2123.
 29. Butler AC, Brown GK, Beck AT, Grisham JR. Assessment of dysfunctional beliefs in borderline personality disorder. *Behav Res Ther* 2002;40:1231-40. Doi:10.1016/s0005-7967(02)00031-1.
 30. Hallquist MN, Pilkonis PA. Refining the phenotype of borderline personality disorder: Diagnostic criteria and beyond. *Personal Disord* 2012;3:228-46. Doi:10.1037/a0027953.
 31. Arntz A, Veen G. Evaluations of Others by Borderline Patients: *J Nerv Ment Dis* 2001;189:513-21. Doi: 0.1097/00005053-200108000-00004.
 32. Barnow S, Stopsack M, Grabe HJ, Meinke C, Spitzer C, Kronmüller K, et al. Interpersonal evaluation bias in borderline personality disorder. *Behav Res Ther* 2009;47:359-65. Doi:10.1016/j.brat.2009.02.003.
 33. Fertuck EA, Grinband J, Stanley B. Facial trust appraisal negatively biased in borderline personality disorder. *Psychiatry Res* 2013;207:195-202. Doi:10.1016/j.psychres.2013.01.004.
 - *34. Masland SR, Hooley JM. When trust does not come easily: Negative emotional information unduly influences trustworthiness appraisals for individuals with borderline personality features. *J Pers Disord* 2019;1-16. Doi:10.1521/pedi_2019_33_404.
- *Replicates the finding that people with BPD have a bias to rate others as untrustworthy. Found that negative affective context may augment this bias, which suggests potential affective mechanisms.**
35. Miano A, Fertuck EA, Arntz A, Stanley B. Rejection sensitivity is a mediator between borderline personality disorder features and facial trust appraisal. *J Pers Disord* 2013;27:442-56. Doi:10.1521/pedi_2013_27_096.
 36. Nicol K, Pope M, Sprengelmeyer R, Young AW, Hall J. Social judgement in borderline personality disorder. *PLoS One* 2013;8. Doi:10.1371/journal.pone.0073440.
 - *37. Richetin J, Poggi A, Ricciardelli P, Fertuck EA, Preti E. The emotional components of rejection sensitivity as a mediator between borderline personality disorder and biased appraisal of trust in faces. *Clin Neuropsychiatry J Treat Eval* 2018;15:200-5.

***Shows that affective, but not cognitive, components of rejection sensitivity mediate the association between borderline personality traits and biased trustworthiness appraisal. This is important for moving toward an understanding of the mechanisms underlying trust bias.**

38. Arntz A, Dietzel R, Dreessen L. Assumptions in borderline personality disorder: Specificity, stability and relationship with etiological factors. *Behav Res Ther* 1999;37:545-57. Doi:10.1016/s0005-7967(98)00152-1.

39. Downey G, Feldman SI. Implications of rejection sensitivity for intimate relationships. *J Pers Soc Psycho* 1996;70:1327-43.

40. Arntz A, Dreessen L, Schouten E, Weertman A. Beliefs in personality disorders: a test with the personality disorder belief questionnaire. *Behav Res Ther* 2004;42:1215-25. Doi:10.1016/j.brat.2003.08.004.

41. Ayduk Ö, Zayas V, Downey G, Cole AB, Shoda Y, Mischel W. Rejection sensitivity and executive control: Joint predictors of borderline personality features. *J Res Pers* 2008;42:151-68. Doi:10.1016/j.jrp.2007.04.002.

42. De Panfilis C, Riva P, Preti E, Cabrino C, Marchesi C. When social inclusion is not enough: Implicit expectations of extreme inclusion in borderline personality disorder. *Personal Disord* 2015;6:301-9. Doi:10.1037/per0000132.

*43. Miano A, Fertuck EA, Roepke S, Dziobek I. Romantic relationship dysfunction in borderline personality disorder—a naturalistic approach to trustworthiness perception. *Personal Disord* 2017;8:281-6. Doi:10.1037/per0000196.

***A rare exception to significant issues with ecological validity in this body of research. This study considers how people with BPD consider the trustworthiness of their real relationship partners, and suggests, like Masland & Hooley (2019) and Richetin et al. (2018), that negative affective processes may provide clues to the mechanisms underlying trust bias.**

*44. King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR. The rupture and repair of cooperation in borderline personality disorder. *Science* 2008;321:806-10. Doi:10.1126/science.1156902.

***Despite its age, this is a seminal study in this area. This was the first use of a behavioral exchange paradigm to study trust processing in BPD, and includes behavioral and neural measures.**

45. Unoka Z, Seres I, Áspán N, Bódi N, Kéri S. Trust game reveals restricted interpersonal transactions in patients with borderline personality disorder. *J Pers Disord* 2009;23:399-409. Doi:10.1521/pedi.2009.23.4.399.

46. Thielmann I, Hilbig BE, Niedtfeld I. Willing to give but not to forgive: Borderline personality features and cooperative behavior. *J Pers Disord* 2014;28:778-95. Doi:10.1521/pedi_2014_28_135.
47. Hepp J, Störkel LM, Kieslich PJ, Schmahl C, Niedtfeld I. Negative evaluation of individuals with borderline personality disorder at zero acquaintance. *Behav Res Ther* 2018;111:84-91. Doi:10.1016/j.brat.2018.09.009.
- *48. Hula A, Vilares I, Lohrenz T, Dayan P, Montague PR. A model of risk and mental state shifts during social interaction. *PLoS Comput Biol* 2018;14. Doi:10.1371/journal.pcbi.1005935.

***This is a computational re-analysis of a seminal trust game study (King-Casas et al., 2008). Shows that people with BPD are less likely to attend to cues that their game partner is irritable, and therefore less likely to alter their behavior to repair trust ruptures.**

49. Preuss N, Brändle LS, Hager OM, Haynes M, Fischbacher U, Hasler G. Inconsistency and social decision making in patients with borderline personality disorder. *Psychiatry Res* 2016;243:115-22. Doi:10.1016/j.psychres.2016.06.017.
- *50. Liebke L, Koppe G, Bungert M, Thome J, Hauschild S, Defiebre N, et al. Difficulties with being socially accepted: An experimental study in borderline personality disorder. *J Abnorm Psychol* 2018;127:670-82. Doi:10.1037/abn0000373.

***Suggests that people with BPD are sensitive to and able to use experiences of partner rejection to adjust their behavior and expectations, but that they react to partner acceptance paradoxically, with less trust and cooperation.**

- *51. Fineberg SK, Leavitt J, Stahl DS, Kronemer S, Landry CD, Alexander-Bloch A, et al. Differential valuation and learning from social and nonsocial cues in borderline personality disorder. *Biol Psychiatry* 2018;84:838-45. Doi:10.1016/j.biopsych.2018.05.020.

***People with BPD are less sensitive to partner volatility. Suggests that people with BPD expect partner betrayal, and are so used to real or perceived signs of partner betrayal that they may fail to notice and respond to these signs.**

52. Fonagy P, Luyten P. A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Dev Psychopathol* 2009;21:1355-81. Doi:10.1017/S0954579409990198.
53. Franzen N, Hagenhoff M, Baer N, Schmidt A, Mier D, Sammer G, et al. Superior 'theory of mind' in borderline personality disorder: An analysis of interaction behavior in a virtual trust game. *Psychiatry Res* 2011;187:224-33. Doi:10.1016/j.psychres.2010.11.012.
54. Meyer-Lindenberg A. Trust me on this. *Science* 2008;321:778-80. Doi:10.1126/science.1162908.

55. Lönqvist J-E, Verkasalo M, Wichardt PC, Walkowitz G. Personality disorder categories as combinations of dimensions: Translating cooperative behavior into the Five-Factor framework. *J Pers Disord* 2012;26:298-304. Doi:10.1521/pedi.2012.26.2.298.

56. De Dreu CKW, Kret ME. Oxytocin conditions intergroup relations through upregulated in-group empathy, cooperation, conformity, and defense. *Biol Psychiatry* 2016;79:165-73. Doi:10.1016/j.biopsych.2015.03.020.

57. Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E. Oxytocin increases trust in humans. *Nature* 2005;435:673-6. Doi:10.1038/nature03701.

58. Mikolajczak M, Pinon N, Lane A, de Timary P, Luminet O. Oxytocin not only increases trust when money is at stake, but also when confidential information is in the balance. *Biol Psychol* 2010;85:182-4. Doi:10.1016/j.biopsycho.2010.05.010.

*59. Bartz J, Simeon D, Hamilton H, Kim S, Crystal S, Braun A, et al. Oxytocin can hinder trust and cooperation in borderline personality disorder. *Soc Cogn Affect Neurosci* 2011;6:556-63. Doi:10.1093/scan/nsq085.

***Despite its age, this study remains highly relevant to our understanding of how OXT influences trust process in BPD. Shows that the effect of OXT depends on attachment.**

*60. Ebert A, Kolb M, Heller J, Edel M-A, Roser P, Brüne M. Modulation of interpersonal trust in borderline personality disorder by intranasal oxytocin and childhood trauma. *Soc Neurosci* 2013;8:305-13. Doi:10.1080/17470919.2013.807301.

***Despite its age, this study remains highly relevant to our understanding of how OXT influences trust process in BPD. Shows not only that OXT can negatively alter trust in BPD, but also that this depends in part on childhood trauma. Also suggests that people with BPD may be using different social cues (i.e., partner attractiveness) to determine trust, which is echoed in more recent work on alterations in social cue responsiveness.**

61. Clifton A, Pilkonis PA, McCarty C. Social network in borderline personality disorder. *J Pers Disord* 2007;21:434-41. Doi:10.1521/pedi.2007.21.4.434.

62. Roberts ID, Krajchich I, Cheavens JS, Campo JV, Way BM. Acetaminophen reduces distrust in individuals with borderline personality disorder features. *Clin Psychol Sci* 2018;6:145-54. Doi:10.1177/2167702617731374.

*63. Fonagy P, Luyten P, Allison E, Campbell C. What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication. *Borderline Personal Disord Emot Dysregul* 2017;4:9. Doi:10.1186/s40479-017-0062-8.

***Discusses the importance of epistemic trust for understanding BPD's development and potential treatments. Conceptualizes BPD as a disorder of social**

communication and low resilience. Describes how epistemic trust develops normally and abnormally. Although this work uses different language than the empirical studies, there is significant overlap and potential for integration.

64. Fonagy P, Luyten P, Allison E. Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *J Pers Disord* 2015;29:575-609. Doi:10.1521/pedi.2015.29.5.575.
65. Fonagy P, Luyten P, Allison E, Campbell C. What we have changed our minds about: Part 1. Borderline personality disorder as a limitation of resilience. *Borderline Personal Disord Emot Dysregul* 2017;4:11. Doi:10.1186/s40479-017-0061-9.
66. Csibra G, Gergely G. Natural pedagogy. *Trends Cogn Sci* 2009;13:148-53. Doi:10.1016/j.tics.2009.01.005.
67. Sperber D, Clément F, Heintz C, Mascaro O, Mercier H, Origgi G, et al. Epistemic vigilance. *Mind Lang* 2010;25:359-93. Doi:10.1111/j.1468-0017.2010.01394.x.
68. Fonagy P, Allison E. The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy* 2014;51:372-80. Doi:10.1037/a0036505.
69. Luyten P, Campbell C, Fonagy P. Borderline personality disorder, complex trauma, and problems with self and identity: A social-communicative approach. *J Pers* 2019;88:88-105. Doi:10.1111/jopy.12483.
70. Bo S, Sharp C, Beck E, Pedersen J, Gondan M, Simonsen E. First empirical evaluation of outcomes for mentalization-based group therapy for adolescents with BPD. *Pers Disord* 2017;8:396-401. Doi:10.1037/per0000210.