

2022

Dynamics of Trust Behaviour in Borderline Personality Disorder

Gamze Abramov

Follow this and additional works at: <https://ro.uow.edu.au/theses1>

University of Wollongong

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material.

Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Dynamics of Trust Behaviour in Borderline Personality Disorder

Gamze Abramov

Supervisors:

Doctor Sebastien Miellet

Senior Professor Frank Deane

This thesis is presented as part of the requirement for the conferral of the degree:

Doctor of Philosophy (Clinical Psychology)

University of Wollongong

School of Psychology

2022

Abstract

Background: Borderline personality disorder (BPD) is associated with difficulty forming trust and maintaining cooperation in trust-based exchanges, but little is known about how the disorder influences the temporal ebb and flow of trust, and what aspects of the disorder might be responsible for these trust patterns. An economic game paradigm, the trust game (TG), was used to examine the trajectories of trust as it formed, dissolved, and restored in response to trust violation and repair. Study 1 examined how these trust patterns varied as a function of the number of BPD traits endorsed. Study 2 investigated whether insecure attachment style, self-protective beliefs, and feelings of rejection moderated the effect of BPD trait count on these trust patterns. Study 3 explored how the social-cognitive reasoning for the decisions made during the TG – before, during, and after the trust violation and repair – varied based on the level of BPD trait count.

Method: In all three studies, young adults ($N=234$) played a 15-round TG in which partner cooperation was varied to signal trust violation and repair, resulting in three phases of trust: formation, dissolution, and restoration. Following the TG, participants were asked to provide the reasoning behind the decisions that they and their partner made. In Study 1, discontinuous growth modelling (DGM) was employed to first model the trajectory of trust during these phases, and the magnitude of the changes in trust in response to the violation and repair, and second, to examine how these trust patterns were influenced by BPD trait count. Study 2 extended the DGM by examining whether the effect of BPD trait count on the levels and trajectory of trust was modified based on attachment style (fearful or preoccupied), endorsing the belief that pre-emptive action should be taken to protect oneself from others, and pre-existing feelings of rejection. In Study 3, 16 of the participants with a high BPD trait count (7-10 traits) and 16 randomly selected, age and gender matched participants with a low BPD

trait count (0-2 traits) were asked to provide reasons to explain their own and their partners behaviour during the TG. These reasons were subsequently categorized and compared for between-group differences across each phase of the game.

Results: In Study 1, BPD trait count was associated with an incongruous pattern of trust behaviour in the form of declining trust when interacting with a new and cooperative partner, and paradoxically, increasing trust following multiple instances of trust violation by that partner. BPD trait count was also associated with trust restoring at a faster rate than it was originally formed. Results from Study 2 suggest that the slower rate of trust formation associated with BPD trait count was accounted for by pre-existing feelings of rejection and self-protective beliefs, each of which predicted a slower rate of trust growth. In contrast, endorsing a preoccupied attachment style was found to temper the trust-negating effect of BPD trait count when trust was forming. The faster rate of trust growth in response to trust violations associated with BPD trait count was no longer significant after self-protective beliefs were accounted for, but the latter's effect on trust during this phase did not reach significance. Study 3 found that during trust formation, the high BPD trait group more frequently attributed negative characteristics to the other player and described their own behaviour as a tit-for-tat strategy despite reasoning that their partner was responding with positive reciprocity. Notably, they articulated these reasons markedly less during the restoration phase compared to the formation phase. Both groups provided similar reasons to describe behaviour during the dissolution phase. Overall, the high BPD trait group was also more likely to say that they did not know why they or the other player made the decisions that they did.

Conclusions: The studies in this thesis adopted an innovative methodological and analytical approach to illustrate both behaviourally and cognitively how relational disturbances may play out in trust-based interpersonal exchanges for those with a high BPD trait count. Interestingly, the studies revealed a paradoxical style of relating where cooperative partner behaviour elicited *less* trusting behaviour and more hostile reasoning by those with high levels of BPD traits. In contrast, these individuals appeared to engage in more trusting behaviours in response to trust violation, but this phenomenon requires further investigation. Interventions which focus on improving the capacity and accuracy of social-cognitive reasoning may address the trust-related interpersonal difficulties associated with BPD. In particular, exploring themes of rejection and the need to self-protect may provide further insight into these incongruous trust behaviours.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervision team. Professor Frank Deane, your consistency, wisdom, guidance, unwavering support, and encouragement carried me through these arduous years. Without you in my corner I'm not sure this PhD could have happened. Thank you for playing such an integral role in my supervision. Dr. Sebastien Miellet, your enthusiasm, passion for research, endless willingness to demystify statistical concepts, and your French irreverence have been much appreciated. I would also like to extend my thanks to Dr. Jason Kautz across the other side of the Pacific. You so generously helped me navigate the world of discontinuous growth modelling and continued to offer your support even when you were moving across the country. I struck gold when I found you!

I would also like to express my deepest appreciation to my parents for the immense support you've given me that has made it possible for me to be both a student and a parent. Mum, whether you were dropping off meals, playing taxi driver to the kids, or pinning ballet buns, you have always tirelessly supported me. Dad, for as long as I can remember I've watched you persevere and teach yourself new and difficult things. Fortunately for me I inherited your tenacity. I would also like to thank my co-parent Yossi. Without your steadfast support my university journey could not have happened. Thank you for all that you did for our children that made it possible for me to study and thank you for sharing your technical and programming expertise. But most of all, thank you for encouraging me to reach for bigger things than I imagined for myself.

There are vast stretches of the PhD journey that are solo, and it is the love and support of friends that help one traverse those stretches. Amalia and Denise, you have made these stretches a little less solitary. Denise, the daily coffee and swims we had at the beach during the early and most difficult years are some of my fondest memories and ensured that each

morning started with a laugh. Amalia, the camaraderie, the loving kindness, the fierce support, the generosity, the laughter, the tears. It has meant so much to me. I look forward to seeing what the next chapter holds for each of us. I would also like to mention a very special person, Richard, who has filled the final six months of this journey with zest, positivity, side-splitting laughter, and a joyous ability to turn a weekend stroll into a choose-your-own-adventure Tarantino-esque memory-making escapade. Thank you for making this part of the journey so very fun.

Now I have come to the most difficult part of the acknowledgments. I would like to dedicate this PhD to my children, Lenin and Ramona. Reflecting on my academic journey, which also includes the undergraduate years, has meant reflecting on your lives. Lenin, you were not yet a year old when I decided to return to university to pursue a career in psychology. The prolific sleeper that you were, I wrote assignments during your marathon naps. We moved from Tokyo to Australia so that I could do my honours year, and there we were joined by you Ramona, unequivocally *not* a napper, but very much a prolific giggler with an impressively sophisticated sense of humour. As I look back at the years that passed that took me from a new mother with a dream to go back to university to a fully-fledged psychologist with her own practice and submitting her PhD, I can't help but see a parallel movie reel of your lives, and like any parent, there is a bittersweet sadness about the parts that I missed because I took on such an ambitious program of study.

Lenin, you are now 16 and a term away from starting your final year of high school. I am so impressed by the young man you are - genuine, thoughtful, level-headed, and so very likeable. Ramona, you are now 12 and in your final year of primary school. I continue to be in awe of your sharp mind, wit, curiosity, and creativity. Ending the story of my PhD makes room for new stories and I look forward to the new stories I get to write with you. I love you both so much!

Certification

I, Gamze Abramov, declare that this thesis, submitted in fulfilment of the requirements for the conferral of the degree Doctor of Philosophy (Clinical Psychology), from the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Gamze Abramov

17 August 2022

Formatting Statement

This thesis is presented in a journal article compilation style format. Each chapter is based on a manuscript that has been published or is under review in a peer-reviewed journal, with the exception of the introduction (Chapter 1) and general discussion (Chapter 5) chapters.

List of Publications and Presentations

Publications/Manuscripts

Abramov, G., Miellet, S., Kautz, J., Grenyer, B. F. S., & Deane, F. P. (2020). The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair. *PLOS ONE*, *15*(7).
<https://doi.org/10.1371/journal.pone.0236170>

Abramov, G., Kautz, J., Miellet, S., & Deane, F. P. (2022). The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count. *Journal of Psychopathology and Behavioral Assessment*. <https://doi.org/10.1007/s10862-022-09965-9>

Abramov, G., Miellet, S., & Deane, F. P. (2022). Social-cognitive reasoning for decisions made in a trust game for individuals with high or low borderline personality disorder trait count. *Manuscript submitted for publication*.

Conference Presentations

Abramov, G., Miellet, S., Bourke, M. E., & Grenyer, B. F. S. (2018, November 2-3). *Social cognition deficits associated with interpersonal trust in borderline personality disorder* [Poster presentation]. 12th International Treatment of Personality Disorders Conference, Wollongong, Australia.

<https://documents.uow.edu.au/content/groups/public/@web/@project-air/documents/doc/uow244419.pdf>

Abramov, G., Miellet, S., Kautz, J., Grenyer, B. F. S., & Deane, F. P. (2019, November 8-9). *The paradoxical decline and growth of trust as a function of borderline personality disorder trait count* [Poster presented]. 13th International Treatment of Personality Disorders Conference, Wollongong, Australia.

<https://documents.uow.edu.au/content/groups/public/@web/@project-air/documents/doc/uow257410.pdf>

Abramov, G., Miellet, S., Kautz, J., Grenyer, B. F. S., & Deane, F. P. (2019, October 15-18). *Trust violation and the paradoxical decline and growth of trust in the borderline personality* [Poster presented]. XVI Congress of the International Society for the Study of Personality Disorders, Vancouver, British Columbia, Canada.

<http://www.isspd.com/isspd-2019-congress>

Statement of Contribution of Others

This statement verifies that the candidate undertook the greater part of the work in the previously stated publications and manuscripts. Under the guidance of her supervisors, Gamze Abramov took primary responsibility for the design, data collection, data analysis, and manuscript preparation of each study. Participant recruitment and data collection was undertaken by Gamze Abramov for Studies 1 (Chapter 2), 2 (Chapter 3), and 3 (Chapter 4). Co-authors Frank Deane and Sebastien Miellet contributed to the thesis as supervisors and provided guidance surrounding the design of each of the three studies and editorial suggestions for each manuscript. Co-author Jason Kautz contributed to the data analysis and preparation of manuscripts for Studies 1 and 2 in this thesis. Co-author Brin Grenyer was involved in initial discussions of the research study plan and preparation of manuscript 1.

Gamze Abramov

Doctor Sebastien Miellet

Senior Professor Frank Deane

Key Abbreviations

BPD	Borderline Personality Disorder
DGM	Discontinuous Growth Modelling
HC(s)	Healthy Control(s)
MU(s)	Monetary Unit(s)
TG	Trust Game

Table of Contents

Dynamics of Trust Behaviour in Borderline Personality Disorder.....	1
Abstract.....	2
Acknowledgments.....	5
Certification	7
Formatting Statement.....	8
List of Publications and Presentations	9
Statement of Contribution of Others.....	11
Key Abbreviations	12
Table of Contents.....	13
List of Figures	16
List of Tables	17
Chapter 1: Literature Review and Research Program Objectives	19
Interpersonal disturbance and borderline personality disorder.....	19
Characteristics of interpersonal difficulties in BPD	20
Conceptual links between trust and BPD.....	21
Attachment theory & BPD.....	21
Mentalization & BPD	23
Empirical links between trust and BPD	25
BPD and trust beliefs	26
BPD and trust appraisal biases.....	26
BPD and trust behaviours	27
Gaps in the Current Understanding of TG Behaviours in BPD.....	28
Trust as a dynamic and multiphasic phenomenon	28
Covariates and contributing factors influencing trust dynamics.....	31
Social-cognitive reasoning and trust behaviours	34

Rationale of Thesis	36
Thesis Aims	37
Outline of Thesis.....	38
Chapter 2: The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair	40
Abstract.....	41
Introduction.....	42
Methods.....	50
Results.....	60
Discussion.....	75
Chapter 3: The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count	83
Abstract.....	84
Introduction.....	85
Method	92
Results.....	98
Discussion.....	106
Chapter 4: Social-cognitive reasoning for decisions made in a trust game for individuals with high or low borderline personality disorder trait count	113
Abstract.....	114
Introduction.....	115
Method	120
Results.....	124
Discussion.....	132
Chapter 5: General Discussion.....	141

Summary of Findings.....	141
Implications.....	144
Empirical and Conceptual Implications.....	144
Clinical Implications.....	154
Methodological Implications.....	156
Limitations.....	158
Future Directions.....	160
Conclusions.....	162
Appendices.....	164
Appendix A: Published Manuscript: The paradoxical decline and growth of trust as a function of borderline personality disorder: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair (Chapter 2).....	165
Appendix B: Published Manuscript: The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count (Chapter 3).....	189
Appendix C: Research Materials (Chapters 2-4).....	204
Ethics Approval.....	205
Brief Study Description on SONA.....	207
Participant Information Statement.....	208
Online Information and Consent Form.....	210
Pre-Game Questions.....	213
Trust Game Instructions and Test Questions.....	221
Trust Game Screenshots: Round 1 Example.....	223
Post-Game Social-Cognitive Reasoning Questions Instructions.....	227
Post-Game Trustee Appraisal.....	232
Participant Debrief.....	233
References.....	234

List of Figures

- Figure 1** Trust Game Structure of a Single Round (*Chapter 2*)
- Figure 2** Number of MUs Sent by the Trustor as a Function of Borderline
Personality Disorder Trait Count (*Chapter 2*)
- Figure 3** Influence of BPD on Trust Growth Moderated by Self Protective Beliefs
(*Chapter 3*)
- Figure 4** Influence of BPD on Trust Growth Moderated by Preoccupied Attachment
Style (*Chapter 3*)

List of Tables

- Table 1** Coding and Interpretation of Change Variables in the Discontinuous Mixed-Effects Growth Models (*Chapter 2*)
- Table 2** Intercorrelations, Means, and Standard Deviations of Study Variables (*Chapter 2*)
- Table 3** Model Comparison Tests for Discontinuous Growth Models, Autocorrelations, and Heteroscedasticity (*Chapter 2*)
- Table 4** Final Discontinuous Growth Models (*Chapter 2*)
- Table 5** Discontinuous Mixed-Effects Growth Models Predicting Trust as a Function of BPD Trait Count after Controlling for Gender and Cognitive Reflective Ability (*Chapter 2*)
- Table 6** Discontinuous Mixed-Effects Growth Models Predicting Trust as a Function of BPD Caseness after Controlling for Gender and Cognitive Reflective Ability (*Chapter 2*)
- Table 7** Coding and Interpretation of Change Variables in the Discontinuous Growth Model (*Chapter 3*)
- Table 8** Intercorrelations, Means, and Standard Deviations of Study Variables (*Chapter 3*)
- Table 9** Discontinuous Mixed-Effects Growth Models Predicting Trust as a Function of BPD Trait Count, Attachment Style, Self-Protective Beliefs, and Feelings of Rejection, after Controlling for Gender and Cognitive Reflective Ability (*Chapter 3*)
- Table 10** Game Round, Trustee Reciprocity and the Associated Phase of Trust for Self and Partner Social-Cognitive Reasoning (*Chapter 4*)

Table 11 Categories of Social-Cognitive Reasoning for Self and Partner Trust Game Behaviours (*Chapter 4*)

Table 12 Social-Cognitive Reasoning for Self and Partner Trust Game Behaviours as Trust Forms, Dissolves, and Restores (*Chapter 4*)

Chapter 1: Literature Review and Research Program Objectives

Interpersonal disturbance and borderline personality disorder

Borderline Personality Disorder (BPD) is a psychiatric condition with a median prevalence rate estimated at between 1.6% to 5.9% of the general population (DSM-5; American Psychiatric Association, 2013), and 15% to 20% of patients in psychiatric hospitals or outpatient clinics (Gunderson, 2011). BPD is a complex and enduring disorder characterised by interpersonal and intrapersonal instability. Disturbed interpersonal functioning has been identified as a core component of BPD in factor analytic studies (Becker et al., 2006; Sanislow et al., 2000, 2002), and is proposed to be a superior diagnostic discriminator of BPD (Gunderson, 2007). Relational disturbances persist across the lifespan of the disorder, and have also been observed in non-clinical populations endorsing high levels of BPD features (Tolpin et al., 2004; Trull et al., 1997). Problematic interpersonal functioning ensues across multiple domains for individuals with BPD or borderline features including with peers (Runions et al., 2021), romantic partners (Lazarus et al., 2019; Navarro-Gomez et al., 2017), and family (Eyden et al., 2016). Long-term prospective studies of BPD reveal that even when symptoms remit, improvement in social functioning is often limited (Gunderson et al., 2011; Zanarini et al., 2010). Critically, interpersonal dysfunction has serious implications for risk of suicide in BPD with interpersonal incidents linked to suicidal ideation (Kaurin et al., 2020), and the reported trigger for three-quarters of both first and most lethal suicide attempts (Brodsky et al., 2006). Consequently, efforts to understand interpersonal disturbance in BPD have increasingly engaged researchers across the last few decades (Jeung & Herpertz, 2014; Lazarus et al., 2014; Lis & Bohus, 2013).

Characteristics of interpersonal difficulties in BPD

Interpersonal dysfunction in borderline pathology most readily manifests as instability in the form of alternating idealization and devaluation of relationships, fear of and desperate efforts to avoid abandonment, marked reactivity of mood, and inappropriate expression of anger (American Psychiatric Association, 2013). Affected individuals are hypersensitive to perceived social rejection, misread other people's intentions and misinterpret neutral social information as negative, have difficulties cooperating and problem solving in social contexts, and display increased hostility and distrust towards others (Lazarus et al., 2014; Lis & Bohus, 2013). BPD features are associated with higher levels of criticism and conflict, and reduced satisfaction and support within social networks (Beeney et al., 2018; Lazarus et al., 2016). Individuals with BPD also report higher levels of emotional reactivity in interpersonal contexts (Dixon-Gordon et al., 2013; Gadassi et al., 2014; Reichenberger et al., 2017), including more intense emotional reactions in response to stressors of an interpersonal nature than community controls (Hepp et al., 2017; Hepp, Lane, et al., 2018). Moreover, they rate their day-to-day interpersonal exchanges as more angry, disagreeable, and sad compared to those with a different or no personality disorder (Stepp et al., 2009), and report greater fluctuations in their perceptions of relationship quality in regards to conflict, criticism, closeness, and support (Lazarus et al., 2020). Finally, individuals with BPD experience greater rupture or termination in their relationships compared to healthy controls (HCs) (Lazarus & Cheavens, 2017), and other patients receiving psychiatric treatment (Clifton et al., 2007).

The relationship between BPD features and interpersonal functioning appears to be bidirectional in that a decrease in perceived relationship quality has been found to be followed by an increase in reported borderline features, and vice versa (Howard et al., 2022).

Therefore, identifying the mechanisms that underpin maladaptive interpersonal functioning, and using this understanding to inform treatment is likely to have a profound effect on the course of BPD. One facet of interpersonal functioning that has both conceptual and empirical links to BPD is that of interpersonal trust (Jeung et al., 2016; Lazarus et al., 2014). Empirical investigation has linked BPD to negatively biased trust beliefs, appraisals, and behaviours (Masland et al., 2020; Poggi et al., 2019). The principal aim of this thesis is to investigate the anomalous interpersonal trust behaviours associated with BPD. Specifically, the studies comprising this thesis focus on improving our understanding of how trust forms, dissolves, and restores as a function of BPD trait count, and the factors that may explain the influence of BPD on these trust dynamics. The next section will describe how interpersonal trust is relevant to several prominent theories of BPD.

Conceptual links between trust and BPD

While this thesis has not set out to test a theory, it is nonetheless informed by the conceptualisation of BPD as a disorder developing out of disturbed early attachment relationships (Fonagy et al., 1995; Gunderson & Lyons-Ruth, 2008; Linehan, 1993) and impairments in the capacity to mentalize (Allen et al., 2008; Bateman & Fonagy, 2004; Fonagy & Luyten, 2009).

Attachment theory & BPD

John Bowlby's attachment theory posits that human infants require a consistent nurturing relationship with at least one or more available, responsive, and sensitive caregivers for healthy psychological development (Bowlby, 1969, 1973, 1980). Mary Ainsworth made critical contributions to our understanding of attachment by developing a measure to classify

the child's attachment to his or her caregiver/s (Ainsworth et al., 1969, 1978; Bowlby et al., 1956). Ideally, buoyed by secure attachments to caregiver/s, the child is proposed to develop working models of self as worthy and lovable, and of others as reliable and responsive (Bowlby, 1973). Main et al. (1985) reconceptualised the individual differences in attachment organization as differences in the internal working models or mental representations of the self in relation to significant others. They further proposed that these models direct our thoughts, feelings, behaviour, attention, and memory in attachment-salient contexts.

Mikulincer (1998) drew parallels between an integral component of many established definitions of interpersonal trust – the belief that a person cares about one's needs and can be depended upon (Rempel et al., 1985) – and the expectations that securely attached individuals hold that significant others will be available, caring, and responsive, and concluded that trust was an integral tenet of secure attachment. This is supported in the empirical literature where a secure attachment style is positively associated with the degree of trust an individual feels in general (Collins & Read, 1990; Mikulincer, 1998), as well as in romantic relationships (Fitzpatrick & Lafontaine, 2017; Givertz et al., 2013; Karantzas et al., 2014; Keelan et al., 1994; Simpson, 1990), in work environments (Camgöz & Karapinar, 2016; Frazier et al., 2015), and in healthcare settings (Holwerda et al., 2013; Klest & Philippon, 2016).

Aetiological models of BPD position it as a disorder that at least in part stems from attachment disturbances arising from suboptimal, adverse, or invalidating caregiver experiences (Fonagy et al., 1995; Gunderson & Lyons-Ruth, 2008; Linehan, 1993). The adult attachment styles most prevalent in BPD are fearful-avoidant attachment and anxious-preoccupied attachment (Agrawal, Gunderson, Holmes, & Lyons-Ruth, 2004). While both styles of insecure attachment share negative beliefs about the self, in considering self in relation to others, individuals with a preoccupied attachment style tend to pursue closeness

and connection while fearfully-attached individuals are defined by the tendency to simultaneously desire and fear becoming close to and dependent on others (Bartholomew & Horowitz, 1991). Gunderson and Lyons-Ruth (2008) note that the interpersonal dysfunction in BPD is characterized by the discordant duality of holding a desire for close relationships and intimacy while at the same time experiencing an intense fear of rejection and abandonment.

Compared to HCs, individuals with BPD hold mental representations of others as more negative and aggressive (Barnow et al., 2009), malevolent (Segal et al., 1992, 1993), untrustworthy, and needing to be protected from (Barazandeh et al., 2016; Bhar et al., 2008). BPD has also been associated with rejection-related interpretation biases (Lobbestael & McNally, 2016). Even when the evidence suggests otherwise, individuals with BPD have a tendency to perceive that they are being rejected (De Panfilis et al., 2015; Domsalla et al., 2014; Renneberg et al., 2012; Staebler et al., 2011; Weinbrecht et al., 2018), and they are more likely to respond with increased feelings of anger and hostility to rejection, perceived or actual (Berenson et al., 2011; Hepp, Lane, et al., 2018; Lazarus et al., 2018; Richmond et al., 2020; Scott et al., 2017). It is possible that the expectations of rejection and betrayal that form the worldview of those with insecure attachment styles hinder the development of trust.

Mentalization & BPD

Drawing on attachment theory, Fonagy and colleagues have conceptualized BPD as a developmental disorder characterized by impairments in the capacity to mentalize, that is, to identify, understand and differentiate the mental states of self and others (Allen et al., 2008; Bateman & Fonagy, 2004; Fonagy & Luyten, 2009). Impairments in mentalization and other

overlapping constructs including social cognition (Sharp et al., 2008), Theory of Mind (Premack & Woodruff, 1978), cognitive empathy (Roepke et al., 2012), metacognition (Dimaggio & Lysaker, 2015; Semerari et al., 2003), and reflective functioning (Chiesa & Fonagy, 2014), are empirically well-established in populations with BPD or borderline features (for reviews/meta-analyses see Bora, 2021; Dinsdale & Crespi, 2013; Jeung & Herpertz, 2014; Lazarus et al., 2014; Mitchell et al., 2014; Németh et al., 2018; Richman & Unoka, 2015; Roepke et al., 2012; Salgado et al., 2020). With few exceptions (see Salgado et al., 2020), research has found that BPD has been associated with diminished capacity to infer the mental states of others in written stories (Baez et al., 2015; Harari et al., 2010; Petersen et al., 2016; Pluta et al., 2018; Zabihzadeh et al., 2017), videos (Andreou et al., 2015; Normann-Eide et al., 2020; Preißler et al., 2010; Ritter et al., 2011; Vaskinn et al., 2015), photographs (Anupama et al., 2018; Dziobek et al., 2011; Unoka et al., 2015; Van Heel et al., 2019; Zegarra-Valdivia & Chino Vilca, 2019), and via self-report questionnaire (Anupama et al., 2018; Badoud et al., 2018; Dimitrijevic et al., 2018; Flasbeck et al., 2017, 2019; Grzegorzewski et al., 2019; Harari et al., 2010; Homan et al., 2017; Martin et al., 2017; New et al., 2012; Perroud et al., 2017; Petersen et al., 2016; Ritter et al., 2011). BPD has also been associated with a reduced ability to convey an understanding of both their own and others' mental states via semistructured interviews (Colle et al., 2019).

In contrast to diminished mentalizing, hypermentalization is defined as the tendency to make inferences about the mental state of others that extend beyond what could be reasonably surmised based on the observable data (Sharp et al., 2013). Hypermentalization has been associated with BPD in adults (Andreou et al., 2015; Normann-Eide et al., 2020; Vaskinn et al., 2015) and adolescents (Cortés-García et al., 2021; Penner et al., 2020; Quek et al., 2019; Sharp et al., 2013; Sharp, Pane, et al., 2011; Somma et al., 2019), although a recent

meta-analytic review suggests it may not be specific to BPD but instead related to psychopathology in general (McLaren et al., 2022).

Impairments in mentalization are associated with interpersonal problems (Berenson et al., 2018; De Meulemeester et al., 2017; Euler et al., 2021). Deficits in the capacity to perceive and interpret social signals to accurately infer mental states may compromise the ability to skilfully navigate interpersonal exchanges, such as those involved in trust processes. Trust behaviours have been linked to social cognition in the form of perspective taking (Fett et al., 2014; Sun et al., 2021), reasoning about the intentions of others (Sharp, Ha, et al., 2011), and theory of mind (Derks et al., 2015). If people with BPD are impaired in their ability to decode the more nuanced signals humans emit to communicate their intentions and motivations, it is possible that impairments in mentalizing have a deleterious effect on the ability to perceive and comprehend social signals that help individuals determine when trust should be given or withdrawn. To date there have not been any studies directly examining mentalization processes and trust in BPD.

Empirical links between trust and BPD

There is a growing body of empirical investigation into the relationship between interpersonal trust and BPD. This body of work is focused on three broad areas of inquiry – beliefs and schemas regarding the trustworthiness of others; appraisals of trustworthiness; and trust behaviours in trust-based interpersonal exchanges.

BPD and trust beliefs

Aetiological models of BPD propose that attachment disturbances originate at least in part from suboptimal, adverse, or invalidating caregiver experiences which give rise to mental representations of others as untrustworthy and rejecting (Fonagy et al., 1995; Gunderson & Lyons-Ruth, 2008; Linehan, 1993). Mistrust/abuse schema or beliefs differentiate BPD patients from HCs (Bach & Farrell, 2018; Butler et al., 2002), suggesting mistrust and expectations of abuse by others are prominent in those with borderline features. People with BPD have been found to endorse the belief that others will betray, exploit and deceive, and that one should pre-emptively act to avoid being rejected, ignored or attacked by others (for review see Barazandeh et al., 2016; Bhar et al., 2008). Finally, compared to HCs, BPD patients hold greater expectations of selfishness from others (Lévay et al., 2021), and are more pessimistic in their predictions about the outcome of an economic exchange game where the risk is interpersonal rather than luck (Unoka et al., 2009).

BPD and trust appraisal biases

BPD is associated with negatively biased appraisals of the trustworthiness of others in photographs or short films (Fertuck et al., 2013; Masland & Hooley, 2020; Miano et al., 2013; Nicol et al., 2013; Richetin et al., 2018). The trust appraisal biases associated with borderline pathology appears to be amplified by priming for negative affect (Masland & Hooley, 2020), and mediated by the anger and anxiety related to rejection sensitivity (Miano et al., 2013; Richetin et al., 2018). Biased trust appraisals among individuals with BPD have also been observed in more ecologically valid contexts in which fears of rejection may be more prominent. For example, in a naturalistic study of couple dyads, Miano and colleagues (2017) found that within the context of intimate relationships, women with BPD compared to

HCs appraised their romantic partners as less trustworthy after discussing a relationship threatening or personally threatening topic. In contrast, appraisal ratings were comparable following discussion of a neutral topic. This body of work suggests that emotional processes influence appraisals of trustworthiness, particularly emotions related to rejection (Miano et al., 2013; Miano, Fertuck, et al., 2017; Richetin et al., 2018).

BPD and trust behaviours

Mayer et al. (1995) defined trust as the “willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party” (p.712). Moreover, they differentiated between trust (or trust intentions) and trust behaviour in that the former involves a “willingness” to assume risk while the latter involves “assuming risk”, that is, trust behaviours are actions taken in which one is vulnerable to the risks posed by misplaced trust in the other party (Mayer et al., 1995 p.712).

Behavioural exchange paradigms such as the trust game (TG: Berg et al., 1995) have been used by researchers to examine trust behaviours in interpersonal exchanges (Johnson & Mislin, 2011), including in BPD populations (Masland et al., 2020; Poggi et al., 2019). TGs typically comprise two players. One player, variously known as the *trustor* or *investor* begins with a monetary endowment and can choose whether to entrust any of it with the other player – the *trustee* – for investment. The amount invested is automatically multiplied by a factor – most often three – before being received by the trustee. The trustee is then given the option to cooperate by sending a sum of their choosing from this tripled amount back to the trustor. Trust is operationalized as the proportion of the original endowment transferred by the

trustor, while trustworthiness is operationalized as the proportion of the tripled endowment returned to the trustor by the trustee. The TG can be played with a human dyad or a programmed agent (computer algorithm), the latter allowing researchers to manipulate responses such as the amount invested or returned. In an iterative (multi-round) game, it is mutually beneficial for both parties to cooperate. Much of the TG literature has reported that BPD is associated with reduced trust (Liebke et al., 2018; Niedtfeld & Kroneisen, 2020; Roberts et al., 2018; Unoka et al., 2009) and reduced trustee cooperation (King-Casas et al., 2008). The following section will highlight the gaps in TG research in BPD populations that the studies in the current thesis aim to address, while also describing the findings in greater detail.

Gaps in the Current Understanding of TG Behaviours in BPD

Based on the TG research undertaken to understand trust issues in borderline populations three areas for further research have been identified. These are: (a) Underutilization of methodological/analytical procedures that acknowledge and effectively exploit the dynamic and multiphasic nature of trust; (b) Lack of examination of theoretically related covariates that may underpin trust behaviours; and (c) Social-cognitive reasoning (mentalization) processes behind trust decisions. Each of these gaps in prior research are addressed in the three studies that comprise this thesis. The rationales for these foci are also outlined below.

Trust as a dynamic and multiphasic phenomenon

Research has supported the presence of at least three phases of trust: formation, dissolution and restoration (Fulmer, 2010; Fulmer & Gelfand, 2013; Korsgaard et al., 2018;

Liebke et al., 2018; Lount et al., 2008; Schilke et al., 2013; Thayer, 2015). Trust formation describes the establishing of trust in a new relationship, trust dissolution describes the decline of trust in response to trust violation, and trust restoration describes the restoring of trust following trust dissolution, in response to efforts by the violating party to repair trust. However, much of the research using the TG across disciplines has failed to utilize methodological and data analytical procedures that sufficiently consider the multiphasic nature of trust (see Korsgaard et al., 2018; Lewicki et al., 2006; Rousseau et al., 1998). There are several exceptions in organizational psychology research (e.g., Fulmer & Gelfand, 2015; Korsgaard et al., 2018), but only one example in the BPD literature (Liebke et al., 2018).

While not explicitly examining the phases of trust, the seminal TG study by King-Casas and colleagues (2008) revealed differentiated trust patterns associated with BPD, which appeared to highlight the process of trust reparation. In their study, human dyads played a 10-round TG in which HC trustors were paired with either HC or BPD trustees (King-Casas et al., 2008). During the first half of the game, trustee behaviour was fairly uniform but in response to faltering cooperation by trustees, trustors began to send less money, ostensibly to signal mistrust. There were remarkable differences in how trustees responded. HC trustees were more likely to respond by returning larger sums of money, which appeared to have the effect of ‘coaxing’ the trustors’ ongoing cooperation. In contrast, BPD trustees were more likely to respond by further reducing the amount repaid, which had a detrimental effect on the size of future investments and more frequently led to a breakdown in cooperation (King-Casas et al., 2008). The underuse of coaxing strategies has also been observed in other contexts. For example, in a study using the Cyberball paradigm (Williams & Jarvis, 2006), a virtual ball-tossing game used to experimentally induce conditions of social inclusion and exclusion, HCs were more likely than BPD patients to respond to partial

social exclusion by increasing ball tosses towards the excluder (Barton et al., 2021). The use of human dyads allowed King-Casas et al. (2008) to reveal that borderline pathology appears to modulate the ebb and flow of trust. Their work reinforced the need for research to explicitly measure how BPD influences trust as it forms, dissolves, and restores. A decade later Liebke et al. (2018) paired human trustors (BPD or HC) with allegedly human trustees, actually a simulated computer algorithm programmed to either return a profit or a loss, to create formation, dissolution, and restoration phases. They found that between the groups, trustors did not differ in how much they invested in either the dissolution or restoration phases, but the BPD trustors made smaller investments than HC trustors during the trust formation phase.

While Liebke et al. (2018) operationalized trust as the mean amount invested during each phase, other trust researchers have used a form of mixed effect modelling called discontinuous growth modelling (DGM: Bliese & Lang, 2016; Singer & Willett, 2003), which offers more sophisticated and nuanced means with which to understand how trust changes. DGM can be used to analyse longitudinal data in which meaningful discontinuities are present such as a trust violation or bid for trust repair, and allows researchers to assess the rate at which trust changes within phases, as well as the magnitude of trust changes between phases due to the discontinuities (Fulmer, 2010; Fulmer & Gelfand, 2015; Korsgaard et al., 2018). By pairing DGM with a TG in which trust ruptures and repairs have been programmed, BPD researchers can better understand how borderline pathology and other individual differences influence the *rate* at which trust forms, dissolves, and restores, and the magnitude of the shifts between these phases. Moreover, DGM also lends itself to more nuanced analyses such as examining whether there are differences in the rate at which trust restores relative to the rate at which it was formed, that is, intraindividual comparisons, and

the influence of variables such as BPD trait count on these within-person trust changes (Bliese & Lang, 2016). The capacity to capture these aspects of trust is particularly promising for research within the BPD sphere since this population acts in ways that jeopardise the development and maintenance of trust (King-Casas et al., 2008; Liebke et al., 2018; Unoka et al., 2009).

Covariates and contributing factors influencing trust dynamics

While the extant research has linked BPD to anomalous trust patterns (King-Casas et al., 2008; Liebke et al., 2018; Niedtfeld & Kroneisen, 2020; Roberts et al., 2018; Unoka et al., 2009), our understanding of contributing or underpinning factors is in its infancy. Attachment theory (Bowlby, 1969, 1973, 1980) provides an organizing framework to understand the trust behaviours observed in BPD, but research examining the impact of attachment style on these behaviours in BPD is lacking. Several research groups have investigated the effect of the neuropeptide oxytocin on trust in individuals with BPD (Bartz, Simeon, et al., 2011; Ebert et al., 2013). Oxytocin is involved in a number of complex social behaviours including parent-child bonding, adult attachments, and pro-social behaviour (for reviews see Bakermans-Kranenburg & Van Ijzendoorn, 2013; Bartz, Zaki, et al., 2011; Galbally et al., 2011; Graustella & MacLeod, 2012; MacDonald & MacDonald, 2010; Szymanska et al., 2017), including trust behaviours and appraisals in healthy adults (De Dreu, 2012; Kosfeld et al., 2005; Theodoridou et al., 2009). Although the oxytocin-trust link has not been conclusively established (Declerck et al., 2020), in people with BPD it appears that oxytocin has a paradoxically trust-hindering effect (Bartz, Simeon, et al., 2011). Bartz and colleagues used an economic game paradigm in which mutual cooperation was incentivised and found that HCs under the oxytocin condition reported greater expectations of partner

trustworthiness and intentions to cooperate compared to those in the placebo condition, although this fell short of significance. However, for individuals with BPD, the oxytocin condition promoted the opposite effects, that is, lower expectations of partner trustworthiness and a greater intention to defect than under the placebo condition (Bartz, Simeon, et al., 2011). A between-group difference on the effect of oxytocin on *actual* trust behaviours was not established. A later TG study found a similarly trust-diminishing effect of oxytocin on inpatients with BPD, and this effect further appeared to be moderated by a history of childhood emotional neglect, with higher levels of neglect predicting smaller investments (Ebert et al., 2013).

Oxytocin also appears to exert a differential effect on trust depending on the *type* of attachment insecurity. After collapsing across diagnosis, attachment anxiety which is elevated in anxiously attached, rejection-sensitive participants (i.e. both preoccupied and fearful attachment styles), appeared to be associated with lower expectations of partner trustworthiness and a greater intention to defect under the oxytocin condition compared to placebo (Bartz, Simeon, et al., 2011). Moreover, oxytocin appeared to modulate actual trust behaviours for anxiously attached, rejection-sensitive participants, but the effect varied depending on levels of attachment avoidance. High anxiety/high avoidance individuals (i.e. fearfully attached) became *less* cooperative under oxytocin, while high anxiety/low avoidance individuals (i.e. preoccupied attached) became *more* cooperative under oxytocin (Bartz, Simeon, et al., 2011). It is possible that the more common attachment presentation in BPD – the paradoxical desire for both closeness and distance - may promote mistrust and diminish willingness to cooperate under conditions of attachment threat. The differentiated responses reinforce the need to consider the effect of borderline pathology on trust behaviours as a function of the type of attachment disturbance.

Other potential contributing factors or covariates conceptually related to attachment include trust-related beliefs and rejection-related affect. It is possible that the borderline propensity for beliefs or schema relating to mistrust and expectations of abuse/betrayal (Barazandeh et al., 2016; Bhar et al., 2008), cast a shadow over interpersonal interactions so that the individual projects malevolent motivations on the other player, and engages in a self-protective style of interaction. Supporting this hypothesis is the finding that in comparison to controls, trustors with BPD both send less money and report more negative predictions of TG outcome even when the trustor is not provided with any feedback regarding whether or how much the trustee has reciprocated (Unoka et al., 2009). Individuals with BPD are also more likely to exhibit rejection-related interpretation biases (Lobbestael & McNally, 2016), and more likely to respond with increased anger and hostility to perceived or actual rejection (Berenson et al., 2011; Hepp, Lane, et al., 2018; Lazarus et al., 2018; Richmond et al., 2020; Scott et al., 2017). Notably, Liebke et al. (2018) found evidence that expectations of rejection in the face of behaviour to the contrary appears to have a diminishing effect on trust. In their study they primed participants with an unrelated social activity in which they experienced either social acceptance or social rejection, after which they played a TG. They found that BPD participants who had been primed with acceptance feedback invested significantly *less* than those primed with rejection feedback. In fact, the more positive the feedback was relative to what was expected by the participant, the smaller the corresponding investment. Liebke et al. (2018) suggested that receiving feedback of social acceptance, especially when rejection is anticipated, resulted in more withholding behaviour among borderline participants. This mistrustful response was also observed in relation to implied partner trustworthiness (Niedtfeld & Kroneisen, 2020). Also using a TG paradigm, they found that compared to HCs, individuals with BPD invested significantly less money with partners rated

a-priori (by an unrelated group of participants) as having trustworthy faces, while there was no between-group difference in money invested in partners rated as having untrustworthy faces (Niedtfeld & Kroneisen, 2020). These findings suggest that actual or implied trustworthiness does not appear to counteract the tendency towards mistrust in those with BPD, and potentially may even amplify it. It is possible that cooperative and non-rejecting behaviour by others flies in the face of expected abuse and betrayal, and therefore may reinforce a more self-protective and less mistrustful stance by the borderline individual.

Social-cognitive reasoning and trust behaviours

Meta-analytic findings suggest BPD is associated with impairments in the process of social-cognitive reasoning (Németh et al., 2018), that is, reasoning about others' mental states in order to explain or predict behaviour (Sabbagh, 2004). King-Casas et al. (2008) augmented their TG with neuroimaging to understand the neurological concomitants of reduced cooperation observed in the BPD trustees. Neuroimaging revealed that unlike the HCs, the BPD group did not exhibit an elevated neural response in the bilateral anterior insula, an area associated with social norm violations, after receiving the smaller investments from the trustor. Rather, for BPD trustees, the neural activity was consistent across both small and large investments. The authors suggested that the trustees with BPD may have had an impaired capacity to perceive the motivation or expectations behind trustor decisions to reduce the size of the investment, perhaps as a result of atypical social norms, and therefore failed to take reparative action via coaxing (King-Casas et al., 2008). A subsequent re-analysis of this data set identified a subgroup of trustees, principally from the BPD group, who exhibited difficulties recognizing and/or adapting to trustor cues of irritability, thereby missing the chance to repair the rupture in trust (Hula et al., 2018). These findings highlight

the need to better understand the social-cognitive reasoning behind trust behaviours associated with BPD. The only TG study in the BPD literature that has attempted to examine social cognition capacities but not reasoning per se found evidence to suggest that BPD may proffer advantages in mental state decoding (Franzen et al., 2011), but their use of the TG differed significantly from the other researchers. Franzen (2011) used single round TG's against multiple partners in comparison to the iterative TG's used with single partners in the other BPD/TG studies (e.g., King-Casas et al., 2008; Liebke et al., 2018; Unoka et al., 2009). It is likely that iterative games place additional demands on social-cognitive reasoning capacity since players need to incorporate new information about partner intentions after each round as opposed to one off encounters.

The TG is well-placed to measure how social signals, represented by behaviours such as the amount of money sent or returned, are responded to. However, without directly asking individuals how they interpreted the behaviours of others or their own behaviour, we cannot be certain that the differences in TG behaviour are due to differences in social-cognitive reasoning. For example, using a TG with boys with externalizing behaviour problems, Sharp and colleagues (2011) asked the boys to provide explanations for their own and their partners decisions. Subsequently they learned that among these boys, the amount of money sent or returned was negatively associated with ascribing unfair intentions to explain their partners behaviour, as well as hostile intentions to explain their own behaviour (Sharp, Ha, et al., 2011). At the time of writing, none of the TG studies within the BPD literature have elicited the social-cognitive reasoning that accompanies the behaviours in the game, reflecting a significant gap and an area of need for research in this field.

Rationale of Thesis

Disturbed interpersonal functioning is a central tenet of BPD (Gunderson, 2007), associated with relationship instability (Clifton et al., 2007; Lazarus & Cheavens, 2017), interpersonal conflict (Beeney et al., 2018; Lazarus et al., 2016), emotional reactivity (Dixon-Gordon et al., 2013; Gadassi et al., 2014; Hepp et al., 2017), and suicidality (Brodsky et al., 2006; Kaurin et al., 2020). Difficulties with relational functioning appear to persist, even when the symptoms of BPD remit (Gunderson et al., 2011; Zanarini et al., 2010). Importantly however, emotionally supportive relationships can positively alter the progression of BPD (Kuhlken et al., 2014; Links & Heslegrave, 2000). Therefore, understanding the mechanisms underpinning the interpersonal dysfunction in this disorder may allow for the development of appropriate interventions that improve relational functioning, and allow these individuals to maintain and benefit from healthier, more emotionally supportive relationships.

One of the likely factors contributing to interpersonal disturbance in BPD has been proposed to be anomalies in trust (Masland et al., 2020; Poggi et al., 2019). BPD is associated with core beliefs that others will betray, abuse, and reject (Barazandeh et al., 2016; Bhar et al., 2008), and that pre-emptive action should be taken to protect oneself (Bhar et al., 2008). Those with BPD make negatively biased appraisals of the trustworthiness of others (Fertuck et al., 2013; Masland & Hooley, 2020; Miano, Fertuck, et al., 2017), and act in ways that compromise the development and maintenance of trust (King-Casas et al., 2008; Liebke et al., 2018; Niedtfeld & Kroneisen, 2020; Roberts et al., 2018; Unoka et al., 2009). This appears to be paradoxically compounded in the face of signals of trustworthiness or actual cooperation (Liebke et al., 2018; Niedtfeld & Kroneisen, 2020). Potential contributors or covariates that may explain the effect of BPD on trust include attachment insecurity, beliefs about the need to self-protect from betrayal by others, feelings of rejection, and social-

cognitive reasoning. Moreover, there is a need to ensure that trust research takes into account the dynamic nature of trust (Korsgaard et al., 2018; Lewicki et al., 2006; Rousseau et al., 1998) and its multiphasic structure (e.g., Fulmer, 2010; Korsgaard et al., 2018; Liebke et al., 2018), particularly given the findings that BPD is associated with greater mistrust during the trust formation phase (Liebke et al., 2018; Unoka et al., 2009), and a decreased tendency to repair ruptures in trust (Hula et al., 2018; King-Casas et al., 2008).

Thesis Aims

This thesis aims to utilize methodological and data analytical procedures to model the potential effects of borderline pathology, operationalized as the number of BPD traits endorsed by the individual, on how trust forms, dissolves, and restores in response to trust violation and repair. It will also assess how these BPD-trust dynamics are modified by self-protective beliefs, attachment style, rejection-related feelings, and reasoning about the behaviour of self and others. Using a multi-round TG and DGM the research begins by modelling how trust forms with a new cooperative partner, dissolves in response to a trust violation, and restores in response to repair bids. Specifically, the following questions will be addressed:

- 1) Does BPD trait count modulate the direction and rate in which trust changes when it is forming, dissolving, or restoring?
- 2) Does BPD trait count modulate the magnitude of the decrease in trust immediately following a trust violation?
- 3) Does BPD trait count modulate the magnitude of the increase in trust immediately following a trust repair?

- 4) Does BPD trait count modulate the rate at which trust dissolves relative to the rate at which it was originally formed?
- 5) Does BPD trait count modulate the rate at which trust restores relative to the rate at which it was originally formed?
- 6) Are the effects of BPD trait count on the rate at which trust forms, dissolves, or restores, and the magnitude of the shifts following the violation or repair, moderated by self-protective beliefs, attachment insecurity, or feelings of rejection?
- 7) What reasoning is associated with individuals with a high BPD trait count compared to those with a low BPD trait count to explain TG decisions made by oneself or by the other player when trust is forming, dissolving, and restoring?

Outline of Thesis

This thesis is presented as three studies, each of which has been prepared as a manuscript for publication in a peer-reviewed journal. Each of these studies will be presented in three corresponding chapters. The manuscripts have been formatted to be consistent with the thesis-style presentation and for the sake of uniformity, including standardizing the terms used. The unmodified published manuscripts for Study 1 and 2 have been included in the Appendices.

Chapter 2 comprises the article titled “The paradoxical decline and growth of trust as a function of borderline personality disorder: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair”, which has been published in *PlosOne* (see Appendix A for published manuscript).

Chapter 3 comprises the article titled “The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count”, which has been published in the *Journal of Psychopathology and Behavioral Assessment* (see Appendix B for published manuscript).

Chapter 4 comprises the article titled “Social-cognitive reasoning for decisions made in a trust game for individuals with high or low borderline personality disorder trait count”, which has been submitted for publication.

Chapter 5 provides a general discussion of the findings from these three studies along with presentation of conceptual, methodological, and clinical implications and recommendations for future research.

Chapter 2: The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair

This chapter has been published in *PlosOne*. The chapter is identical to the published manuscript with several exceptions. To ensure uniformity in formatting across the thesis the term “investor” has been replaced by the term “trustor”. To improve readability, Table S2 which was included as a supporting file in the published manuscript was included in the chapter (Table 6). A number of other minor formatting changes have been made in the service of overall uniformity and readability.

Abramov, G., Miellet, S., Kautz, J., Grenyer, B. F. S., & Deane, F. P. (2020). The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair. *PLOS ONE*, *15*(7), e0236170.
<https://doi.org/10.1371/journal.pone.0236170>

Abstract

Borderline personality disorder (BPD) is associated with paradoxical trust cognitions and behaviours. While BPD is associated with difficulty forming trust and maintaining cooperation in trust-based exchanges, design and analytical methodology best suited to reveal the temporal ebb and flow of trust have been underutilized. We used an economic game to examine the trajectories of trust as it forms, dissolves, and restores in response to trust violation and repair, and to explain how these vary as a function of borderline pathology. Young adults ($N=234$) played a 15-round trust game in which partner trustworthiness was varied to create three phases: trust formation, trust violation, and trust restoration. Discontinuous growth modelling was employed to capture the trends in trust over time and their relationship with BPD trait count. BPD trait count was associated with an incongruous pattern of trust behaviour in the form of declining trust when interacting with a new and cooperative partner, and paradoxically, increasing trust following multiple instances of trust violation by that partner. BPD trait count was also associated with trust restoring at a faster rate than it was originally formed. By adopting a methodology that recognizes the dynamic nature of trust, this study illustrated at a micro level how relational disturbances may be produced and maintained in those with a moderate to high BPD trait count. Further investigation of the factors and processes that underlie these incongruous trust dynamics is recommended.

Introduction

Borderline Personality Disorder (BPD) is a complex and often enduring disorder with a prevalence rate of 1-2% in the community (Sansone & Sansone, 2011), and 15 to 20% among patients in psychiatric hospitals or outpatient clinics (Gunderson, 2011). Disturbed interpersonal functioning has been identified as a core component of BPD in factor analytic studies (Becker et al., 2006; Sanislow et al., 2000), and is one of the strongest diagnostic discriminators of the disorder (Zanarini et al., 1990). Long-term prospective studies show that even when symptoms remit, improvement in social functioning is limited, with profound and persistent relational impairments (Gunderson et al., 2011; Zanarini et al., 2010). Social network analysis has revealed that individuals with BPD experience a greater number of conflicted relationships and are typically ‘cut off’ from more people in their networks (Clifton et al., 2007; Lazarus et al., 2016; Lazarus & Cheavens, 2017), suggesting their relationships are marred by rupture and, potentially, a lack of reconciliation. Since the maintenance of stable partner relationships is associated with recovery (Zanarini et al., 2015), it is essential to identify the relational dynamics that contribute to the instability and breakdown of relationships.

Impairments in the capacity to trust has been proposed as a key factor in understanding the relational disturbances of BPD (Fonagy et al., 2015). Believing that others will betray, exploit and deceive is characteristic of the disorder (Barazandeh et al., 2016; Bhar et al., 2008). A systematic review on early maladaptive schemas found the mistrust/abuse schema was one of the most highly endorsed among people with BPD traits or a diagnosis (Barazandeh et al., 2016). This suggests that these individuals perceive the world and others as malevolent, hostile, and dangerous. Accordingly, experimental studies using economic exchange games have found that individuals with BPD often behave in ways that

compromise the formation and maintenance of trust, particularly once trust has been ruptured (King-Casas et al., 2008; Liebke et al., 2018; Unoka et al., 2009). Among non-clinical populations, diminished interpersonal trust within intimate relationships is associated with diminished perceptions of relationship quality (Givertz et al., 2013), and that fluctuate markedly within short periods of time (L. Campbell et al., 2010). Moreover, breach of trust is associated with relationship dissatisfaction, and is considered to be a relevant factor when deliberating whether to end a relationship (Joel et al., 2017). As such, the way that individuals navigate the vicissitudes in trust is likely to impact the quality, stability, and longevity of relationships, and is therefore relevant to improving interpersonal functioning in BPD. Using an economic game paradigm, the current study seeks to examine how the presence of BPD traits modifies interpersonal trust processes, including how trust changes in response to violation and repair.

The Trust Game (TG)

The experimental paradigm most frequently used to examine trust behaviours is the trust game (TG: Berg et al., 1995). Typically, in a two-player game, one person – the *trustor* – is allocated an endowment (e.g., \$10) and can choose whether to entrust any of it with the other player – the *trustee* – for investment. The amount invested is automatically multiplied by a factor – most often three – before being received by the trustee. The trustee can then reciprocate if they so desire, by sending the trustor a sum of their choosing from this tripled amount. Trust is operationalized as the proportion of the original endowment transferred by the trustor. The TG can be played with a human dyad or a programmed agent, the latter allowing researchers to manipulate responses such as the magnitude of the sum repaid, to investigate how trust changes over time and in response to discrete events.

Interpersonal trust as a dynamic phenomenon

While the TG is an ideal vehicle for examining trust dynamically, the methodological and data analytical procedures best suited to reveal the dyadic ebb and flow of trust and cooperation across repeated interactions have been underutilized (see Korsgaard et al., 2018; Lewicki et al., 2006; Rousseau et al., 1998), with a few exceptions within the organizational psychology sphere (e.g., Fulmer & Gelfand, 2015; Korsgaard et al., 2018).

Trust as a temporal phenomenon has been conceptualized as comprising at least three distinct phases: formation, dissolution, and restoration (Fulmer, 2010; Fulmer & Gelfand, 2013; Korsgaard et al., 2018; Liebke et al., 2018; Lount et al., 2008; Schilke et al., 2013; Thayer, 2015). Trust formation refers to the development of trust in a new relationship; trust dissolution refers to the decline of trust in response to a violation of one's trust by another party; and trust restoration refers to the rebuilding of trust subsequent to trust dissolution, in response to reparative attempts by the offending party. Given that individuals with BPD behave in ways that jeopardise trust development and maintenance (King-Casas et al., 2008; Liebke et al., 2018; Unoka et al., 2009), it is important to conduct research that captures its temporal and mutable nature.

Trust and BPD

In the BPD literature, trust has not typically been treated as a dynamic and multiphasic construct. Rather, it has been measured at a single point in time (Ebert et al., 2013), aggregated across multiple time points (Franzen et al., 2011; Liebke et al., 2018), or restricted to focus on a single phase of trust (Unoka et al., 2009). To the authors' knowledge, only one study has accounted for trust dissolution and trust restoration by manipulating trustee reciprocity to differentiate between three phases: cooperative reciprocity, where the trustee returned a profit across five consecutive rounds; trust rupture, where the trustee kept

the entire investment; and trust repair, where the trustee behaved cooperatively following rupture (Liebke et al., 2018). Trust was operationalized as the amount transferred by the trustor to the trustee for investment, averaged across the aggregated rounds of each phase. Liebke et al. (2018) found no difference between BPD patients and healthy controls (HC) in the average amount invested during the rupture or repair phases but found evidence that patients may transfer less during the cooperative phases, what might loosely be defined as the trust formation phase.

The latter findings were consistent with a 5-round TG study in which feedback on trustee reciprocity was withheld (Unoka et al., 2009). Results indicated that average levels of trust were lower for patients with BPD than patients with major depressive disorder (MDD) or HCs. Linear trend analyses further revealed that while both control groups progressively increased their investments over the five rounds, growth was not observed for the BPD group. Together these findings suggest that individuals with BPD have a greater propensity towards mistrust when interacting with a new partner, even in the absence of investment loss. The temporal nature of trust in BPD was also examined in a study focusing on trustee rather than trustor behaviour. King-Casas et al. (2008) used human dyads to elegantly quantify the process of trustee cooperation in negotiating trust rupture and repair by pairing HC trustors with either HC or BPD trustees in a 10-round TG. They discovered that in response to declining investments by the trustor - a signal of diminishing trust - healthy trustees increased their reciprocity over subsequent rounds, effectively 'coaxing' the investor to send larger investments. In contrast, BPD trustees were more likely to respond by reducing reciprocity further or keeping the entire investment, resulting in increased incidences of breakdown in cooperation with trustors sending even less or nothing in subsequent rounds. They also found differential neurological activity in the bilateral anterior insula between BPD and HC trustees

following receipt of a smaller investment. This area is associated with defection in social exchanges following norm violations (for reviews see Rilling & Sanfey, 2011; Zinchenko & Arsalidou, 2018), leading King-Casas and colleagues' to propose that atypical social norms may underlie the reduced incidence of coaxing behaviours among BPD trustees, by way of a failure to recognize and/or respond appropriately to the social signals of reduced trust.

Although King-Casas and colleagues' study measured trustee cooperation rather than trust, their findings compliment the BPD trust literature to suggest that individuals with BPD compromise the maintenance of trust by not acting to preserve trust when a rupture has occurred (King-Casas et al., 2008). Considered together, the extant research highlights the anomalous relationship between BPD and interpersonal trust, but also exposes gaps in the research, particularly in understanding how trust dissolves and is rebuilt in relation to BPD. Considering this, the current study has paired the TG with appropriate methodological and data analytical procedures better suited to address the dynamic and multiphasic nature of trust.

Measuring change with discontinuous growth modelling (DGM)

In recent years, researchers in the organizational psychology field have advanced the research into trust as a dynamic process by employing longitudinal designs and data analytic approaches suitable to capture how trust unfolds across the various phases of trust (Fulmer, 2010; Fulmer & Gelfand, 2015; Korsgaard et al., 2018). These studies have utilized discontinuous growth modelling (DGM: Bliese & Lang, 2016; Singer & Willett, 2003), a derivative of mixed effects modelling that can model longitudinal data whilst accounting for discontinuities in the data such as an experimenter-induced trust violation event. This

methodological and analytical pairing has enabled examination into the individual differences and higher-level factors that influence how trust forms, dissolves, and restores.

In addition to allowing researchers to measure changes in trust over time and in response to specific events, the change parameters can be coded to measure either *absolute* change in trust or the *relative* intraindividual fluctuations in trust. While using relative coding is the standard in DGM (e.g., Bryk & Raudenbush, 2002; Singer & Willett, 2003), Bliese and Lang (2016) recommend that modelling both types of change may better inform theory development and practical applications. For example, how quickly trust is rebuilt in *absolute* terms can be determined by comparing the trajectory of trust during the trust restoration phase to zero, and further test whether this trajectory is modified by the presence of BPD symptoms. On the other hand, how quickly trust is rebuilt *relative* to how quickly it was initially built can be determined by comparing the trajectory of trust during restoration to the trajectory of trust before the violation took place. As a second step, it is then possible to determine whether BPD moderates the *relative* difference in the trajectories. Examining relative change is especially important if there is a notable linear trend during the period before a discontinuity. The comparatively slower rate of trust growth observed in BPD patients in the initial stages of an exchange (Unoka et al., 2009) supports the inclusion of relative coding.

Current study and aims

The current study uses a multi-round TG and DGM to examine how BPD impacts interpersonal trust processes, including the development of trust with a new partner, how trust dissolves in response to trust violation, and the restoration of trust in response to trust repair. Based on the methodology adopted in previous works (Fulmer, 2010; Fulmer & Gelfand, 2015; Korsgaard et al., 2018), a 15-round TG is used with modified reciprocity to precipitate

a distinct trust violation, before resuming a pre-violation rate of reciprocity to precipitate a trust repair. These variations in reciprocity are intended to elicit the three trust phases of formation, dissolution, and restoration. This allows five change parameters to be modelled: three parameters representing the rate of trust growth within each of the phases and two parameters representing the change in the level of trust between the phases.

The current study primarily aims to examine whether the number of BPD traits reported modifies, (a) how much trust decreases immediately following the violation (dissolution transition), and how much trust increases immediately following the repair (restoration transition), and (b) the direction and rate in which trust changes in each phase (formation slope, dissolution slope, and restoration slope). Based on previous findings that BPD is associated with increased mistrust in the early stages of a social exchange (Liebke et al., 2018; Unoka et al., 2009), and a reduced tendency to utilize trust reparative behaviours to maintain cooperation in a social exchange (King-Casas et al., 2008), it is hypothesized that BPD trait count will be associated with a more pronounced decrease in trust after the initial instance of violation, a less pronounced increase in trust after the initial instance of repair, and a deleterious effect on the trajectories of trust within each of the formation, dissolution, and restoration phases.

Our secondary aims are to examine the fluctuations in trust at an intraindividual level and to determine whether BPD trait count modifies the overall pattern of these fluctuations. In order to determine how the number of BPD traits moderates the rate of trust growth during the dissolution and restoration phases, we switch our focus from *absolute* differences as described in the primary aim, to *relative* differences. Specifically, the study will examine (a) the rate that trust dissolves *relative* to how quickly it was originally built (i.e., dissolution

phase vs formation phase); (b) the rate trust is rebuilt *relative* to how quickly it dissolved (i.e., restoration phase vs dissolution phase); and (c) the rate trust is rebuilt *relative* to how quickly it was originally built (i.e., restoration phase vs formation phase).

The uncooperative behaviour observed among the trustees with BPD in King-Casas et al. (2008) did not appear until the latter half of the TG, suggesting their ability to maintain a trust relationship deteriorated over time. Given that BPD traits are expected to be associated with a slower rate of trust growth during the formation phase, it is hypothesized that even when this initial pattern of trust change is considered, trust growth during the latter trust phases will be even slower for those with a high number of BPD traits. In other words, as BPD trait count increases, the trajectory of monetary units transferred during dissolution and/or restoration in comparison to the trajectory of transfers made during formation, will be more negative.

The final hypothesis refers to changes in the rate and direction of change during restoration relative to dissolution. Based on King-Casas et al's (2008) finding that trustees with BPD were less likely to coax by increasing reciprocity in order to maintain cooperation, it is possible that in a similar vein, trustors with high levels of BPD traits may not respond as favourably to the increase in reciprocity levels in regards to increasing their rate of trust growth. It is hypothesized that even after taking into account the rate of trust dissolving in the dissolution phase, BPD trait count will be associated with a slower rate of trust growth during restoration.

Cognitive reflection – the act of problem solving by engaging in conscious deliberation and suppressing intuitive/impulsive responding (Frederick, 2005) – has been

positively associated with trust in TG's (Corgnet et al., 2016). Additionally, as BPD is associated with impairments in executive functioning (see G. McClure et al., 2016; Ruocco, 2005), altered decision making (see Paret et al., 2017), deficits in social problem-solving (see Lazarus et al., 2014), and impulsivity in interpersonal contexts (Berenson et al., 2016), including a measure to control for cognitive reflection may be warranted. In regards to gender, a meta-analysis found that female trustors invest significantly less than males in the TG (Johnson & Mislin, 2011). As BPD is diagnosed at a rate of 3:1 in females compared to males (APA, 2013), accounting for the effects of gender when examining the influence of BPD traits on trust is justified. Consequently, in both the absolute and relative models, the main effects of cognitive reflective ability and gender will be controlled.

Methods

The study was approved by the University of Wollongong ethics committee HE2017/253. All participants provided written informed consent.

Participant

Participants were undergraduate students from a large Australian university who elected to take part in a psychology research participation program in exchange for course credit. As part of an additional study, after playing the TG, participants were asked to describe their own and their partner's intentions in reference to specific transactions made. Participants whose responses indicated scepticism that their partner was human (and not a computer algorithm) were dropped from the final analyses ($n = 7$), leaving a final sample of 234 (64% female; $M_{age} = 20.87$, $SD_{age} = 5.66$). The recruitment information advised participants that the online study was looking at the relationship between economic decision

making and various personality variables and would involve questionnaires and playing an investment game.

Trust game (TG) protocol

The current study used a multi-round TG comprising of 15 sequential rounds played with the same partner. At the beginning of each round the trustor was allocated 100 monetary units (MU) by the experimenter and given the option to send the trustee any proportion from 0-100% for investment. The amount sent was automatically tripled, and the trustee could repay any proportion from 0-100% of the tripled investment to the trustor. During rounds 1-4 and 8-14 (inclusive), trustee repayments were randomized to fall between 34-44% of the tripled investment, providing the trustor with a return the equivalent of the original investment plus up to 32% profit (range 0-32%). A trustee reciprocity range of 34-44% was selected to ensure repayments would be comparable to the reciprocity ratios observed in human trustees in previous research (Johnson & Mislin, 2011; King-Casas et al., 2008). During rounds 5-7 (inclusive), trustee repayments were randomized to fall between 0-10% of the tripled investment, providing the trustor with a return the equivalent of losing from 70% to 100% of their original investment (due to rounding to the nearest whole number, investments of 1 MU did not incur a loss). This loss was designed to induce a trust violation, and it was repeated over three consecutive rounds to reinforce the participants' perception that the violation was deliberate and not construed as a mistake (Fulmer, 2010; Sitkin & Roth, 1993). Based on this repayment schedule, investments can be divided into three distinct trust phases: formation (rounds 1-5), dissolution (rounds 6-8), and restoration (rounds 9-15). Figure 1 illustrates the process of a single exchange and the rates of trustee reciprocity preceding each event. Following each round, the participant was provided with a summary indicating how much they invested, how much was repaid, and the final sum each party

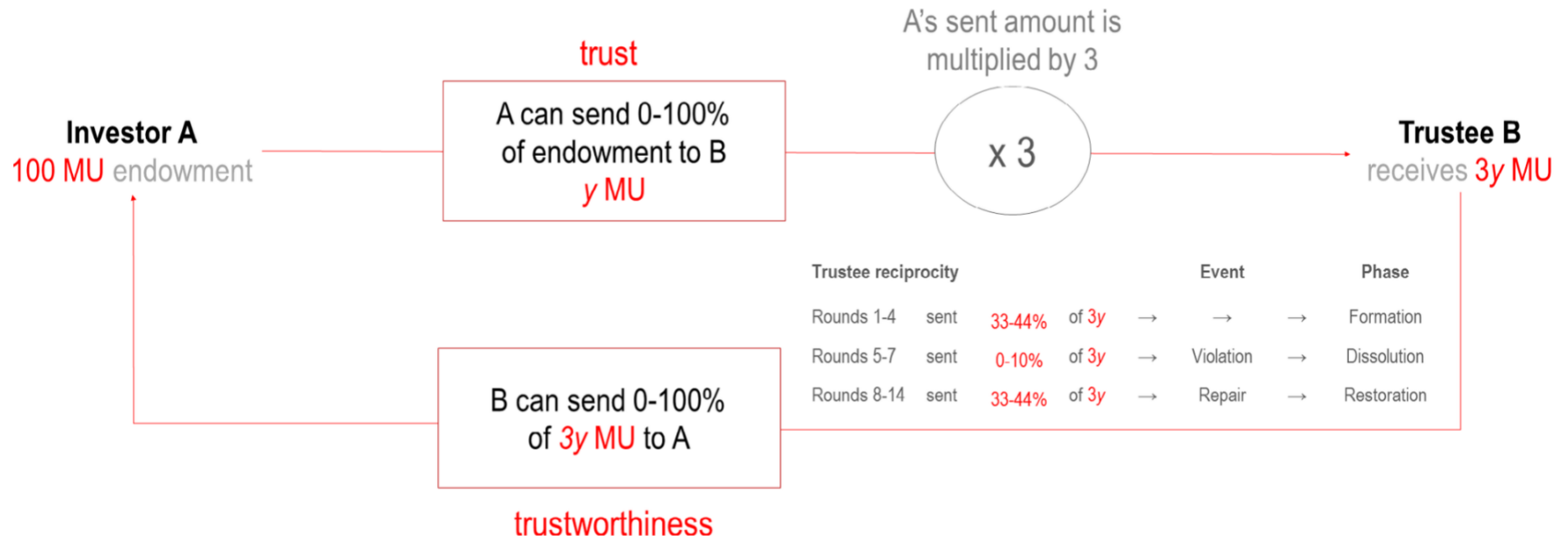
ended up with for that round. As participants saw the summary at the end of each round, trust rates lag trustee reciprocity rates by 1 round.

Procedure

Participants registered and completed the study online during the study period. After completing a series of questionnaires, participants were given instructions on how to play the TG and were required to correctly answer three test questions to demonstrate their understanding of the game before being permitted to proceed. Despite being told roles would be allocated randomly, participants were all assigned the role of trustor. They were also informed that they would be paired with another study participant from a participating academic institution, when in fact the other ‘player’ was a computer programmed trustee. This form of deception has been used in most of the experiments using the same protocol in a BPD population (e.g., Franzen et al., 2011; Unoka et al., 2009). The use of a computerized agent rather than a human confederate allows standardization of trustee responses in terms of reciprocity levels and response time. Participants were not advised how many rounds they would play as previous research has shown that the defection rate increases when people know when a social exchange will end (Murnighan, 1981). Participants were not offered a monetary incentive to participate but at the conclusion of the study, they were automatically allocated course credit.

Figure 1

Trust Game Structure of a Single Round



Note. At the start of each round, trustors were allocated 100 monetary units (MUs) from which they could send any amount to the trustee for investment. After receiving the transferred amount multiplied by a factor of 3 ($3y$ MU), the trustee could then return any amount to the trustor. Trustee reciprocity was randomized to fall between 33-44% and 0-10% to induce violation and repair, respectively.

Measures

Trust

The number of MU that participants entrusted to their partner in each round, ranging from 0 to 100, represents a single behavioural measure of trust. Each participant provided 15 trust measures in total.

BPD trait count

The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD; Zanarini et al., 2003) is a 10-item self-report screening instrument for BPD, with demonstrated internal consistency, validity, test-retest reliability, and in individuals aged 18 to 25, excellent sensitivity (.90) and specificity (.93) for the diagnosis of DSM-IV BPD (APA, 1994). The MSI-BPD has satisfactory internal consistency (Cronbach's $\alpha = .74$, $N = 200$) and good test and retest reliability (Spearman's $\rho = .72$) (Zanarini et al., 2003). In the current study the MSI-BPD was used as a measure of BPD trait count and provided a score ranging from 0 to 10.

In our sample the MSI-BPD demonstrated very good internal consistency (Cronbach's $\alpha = .84$, $N = 234$). The MSI-BPD additionally had significant moderate positive correlations with the Mental Health Inventory (MHI-5; Berwick et al., 1991), a measure of general mental health and quality of life ($r = .54$, $p < .001$), and the Standardized Assessment of Personality – Abbreviated Scale Self-Report (SAPAS-SR; Germans et al., 2008), a measure of general personality psychopathology ($r = .54$, $p < .001$). In the current sample, 24% reported 3 or more BPD traits, a level of symptomatology that is considered to reflect the presence of a notable level of borderline pathology (CLPS: Gunderson et al., 2011).

Finally, using the MSI-BPD conservative cut-off of 7/10, 16 participants met caseness for BPD.

Cognitive reflective ability

The Cognitive Reflection Test (CRT; Frederick, 2005) is a three-item measure of the willingness to engage in deliberation during a cognitive task. Each item is a deceptively simple mathematical problem in which there is an intuitive but incorrect answer. The CRT aims to measure the ability or disposition to resist responding impulsively by assessing the number of questions answered correctly. Participants are required to answer each question by typing the correct response in digits. Correct responses receive a score of 1 with all other responses scored 0, resulting in a total summed CRT score ranging from 0 – 3. The CRT has a moderate positive correlation with measures of intelligence and is correlated highly with various measures of decision-making indices (Frederick, 2005). In the current study 48% of participants scored 0, 18% scored 1, 17% scored 2, and 17% scored 3, with an overall sample mean of 1.03, which falls within the range of CRT scores collected from other academic institutions (see Frederick, 2005).

Post-game trustee appraisals

After the game participants rated the trustee on fairness (*‘Did the other player play fair?’*) and trustworthiness (*‘Is the other player trustworthy?’*) on 5-point Likert-type scales (0 = “Not at all”; 4 = “Absolutely”).

Data Analyses

The NLME package (Pinheiro et al., 2019), included in the open source software R (R Development Core Team, 2018), was used to conduct DGM analyses (see Bliese & Lang,

2016; Singer & Willett, 2003). We tested two-level mixed-effects models, with investment occasions (rounds) at Level 1 nested within individuals at Level 2. Level 1 parameters were coded based on the framework recommended by Singer and Willet (2003) and Bliese and Lang (2016), to examine changes in the average level of trust between and growth within each of the formation (pre-violation), dissolution, and restoration phases. Five change variables were created to examine *absolute* and *relative* change. According to Bliese and Lang (2016), the coding and combination of these change variables into a matrix allows for the regression coefficients to describe the change in the dependent variable in different ways. As we were interested in how individuals respond first to a trust violation and then to a trust restoration, we coded the change variables such that the coefficients reflect the previous stage as the baseline for interpretation. Specifically, the change variables coded for the dissolution phase (DT and DS) use the formation phase as a baseline and change variables coded for the restoration phase (RT and RS) use the dissolution phase as a baseline. The coding and interpretation for each change variable are presented in Tables 1a-b for absolute and relative change, respectively. This coding allowed for easy interpretation of how individuals responded during the three phases of formation, dissolution, and restoration. Finally, an alternative coding system was used to reflect relative change using the formation phase as a baseline (see Table 1c).

We initially examined Level 1 change by including only Level 1 predictors in a series of models to calculate a basic DGM. Next, we examined the Level 2 model by including BPD trait count as a Level 2 predictor to account for differences in Level 1 change, while also controlling for the main effects of gender and cognitive ability. Snijders and Bosker (1999) maintain that as a result of reduced parameter reliability in multilevel analysis, the power to detect cross-level interactions may be insufficient (p. 94).

Table 1

Coding and Interpretation of Change Variables in the Discontinuous Mixed-Effects Growth Models^a

a) Absolute Change

Change Variable	Formation					Dissolution			Restoration						Interpretation of Associated Parameter Estimates in the DGM	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14		15
Rounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
TIME _{ti}	0	1	2	3	4	4	4	4	4	4	4	4	4	4	4	Linear change of MUs transferred in the formation phase (π_{1i})
DT _{ti}	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	Difference in level of MUs transferred immediately following the trust violation (π_{2i}), Round 6 vs Round 5
DS _{ti}	0	0	0	0	0	0	1	2	2	2	2	2	2	2	2	Linear change of MUs transferred in the dissolution phase
RT _{ti}	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	Difference in level of MUs transferred immediately following the trust repair (π_{4i}), Round 9 vs Round 8
RS _{ti}	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	Linear change of MUs transferred in the restoration phase

b) Relative Change (relative to preceding phase)

Change Variable	Formation					Dissolution			Restoration						Interpretation of Associated Parameter Estimates in the DGM	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14		15
Rounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
TIME _{ti}	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Linear change of MUs transferred in the formation phase (π_{1i}), also referred to as the pre-violation period
DT _{ti}	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	Difference in MUs transferred immediately following the trust violation (π_{2i}) relative to the predicted transfer amount based on the

DS_{it}	0	0	0	0	0	0	1	2	3	4	5	6	7	8	9	formation phase (i.e., the expected MUs at Round 6 in the absence of trust violation) Linear change of MUs transferred in the dissolution phase relative to the formation phase (i.e., the pre-violation period) (π_{3i})
RT_{it}	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	Difference in MUs transferred immediately following the trust repair (π_{4i}) relative to the predicted transfer amount based on the dissolution phase (i.e., the expected MUs at Round 9 in the absence of trust repair)
RS_{it}	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	Linear change of MUs transferred in the restoration phase relative to the dissolution phase (i.e., the pre-repair period) (π_{5i})

c) Relative Change (relative to formation phase)

Change Variable	Formation					Dissolution			Restoration							Interpretation of Associated Parameter Estimates in the DGM
Rounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
$TIME_{it}$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Linear change in MUs transferred in the formation phase (π_{1i}), also referred to as the pre-violation period
DT_{it}	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	Difference in MUs transferred immediately following the trust violation (π_{2i}) relative to the predicted transfer

To account for this, a criterion level of $p < .10$ was used for all cross-level interactions effects, and $p < .05$ for all other effects (see also Niessen & Jimmieson, 2016; Yeo & Neal, 2004), and all tests conducted were two-tailed. We tested a model examining *absolute* change to address our first research aim followed by models examining *relative* change to address our second research aim. For the mixed-effects analyses, all level 2 predictors were z-standardized and centered at the sample mean.

Results

Descriptive data and intercorrelations

The dataset for the current study can be accessed in S1 Dataset.¹ Table 2 presents the means, standard deviations, and intercorrelations for BPD trait count, MUs transferred within each trust phase, partner appraisals, and cognitive reflection.

BPD trait count had a significant negative correlation with MUs transferred during formation but was not found to be significantly associated with MUs transferred during the dissolution or restoration phases. BPD trait count was not significantly associated with appraisals of trustworthiness or fairness. However, trustworthiness appraisals were positively associated with the number of MUs transferred during all three trust phases while fairness appraisals were positively associated with the amount transferred during the dissolution phase only.

¹ S1 can be accessed via the online version of this published paper.

Table 2*Intercorrelations, Means, and Standard Deviations of Study Variables*

	1	2	3	4	5	6	7
1. BPD	-						
2. Formation	-0.18**	-					
3. Dissolution	-0.01	0.34***	-				
4. Restoration	-0.08	0.44***	0.55***	-			
5. Trustworthiness	-0.10	0.19**	0.16*	0.15*	-		
6. Fairness	-0.10	0.04	0.15*	0.07	0.49***	-	
7. Cognitive reflective test	-0.11 [†]	0.23***	0.03	0.13 [†]	0.03	0.03	-
Mean	1.63	46.16	24.54	38.4	2.12	2.71	1.03
SD	2.34	18.95	19.14	23.29	1.08	1.18	1.15

BPD traits (0-10). Formation/Dissolution/Restoration (0-100). Trustworthiness/Fairness appraisal (1-5). Cognitive reflective test (0-3).

[†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$. $n=234$ participants. Spearman correlation.

To examine whether there were systematic differences in trustee reciprocity rates we conducted both correlation and ANOVA analyses so that we could treat BPD trait count as a continuous and categorical variable. We created three BPD categories based on number of traits endorsed on the MSI-BPD: Low (0-2), Moderate (3-6), and High (7-10). Due to significant discrepancies in sample size ($N = 177/41/16$, respectively) and evidence of non-normality and heterogeneity of variance in some distributions, we elected to run a non-parametric ANOVA analysis. Both, the correlation approach and the Kruskal–Wallis H Test did not show any significant difference in trustee reciprocity between BPD categories. Cognitive functioning had a marginally significant negative correlation with BPD trait count and was positively associated with MUs transferred during formation and dissolution (see Table 1). T-tests suggest that transfers made during formation differ by gender ($p < .01$), with

females ($M = 43.75$, $SD = 17.77$) transferring fewer MUs than males ($M = 50.38$, $SD = 20.27$), and during restoration ($p < .001$), with females ($M = 34.39$, $SD = 22.12$) transferring fewer MUs than males ($M = 45.43$, $SD = 23.75$). Females ($M = 23.56$, $SD = 18.42$) and males ($M = 26.24$, $SD = 20.35$) transferred comparable MUs during dissolution ($p = ns$). T-tests also indicated a marginally significant difference in BPD trait count by gender ($p = .06$), with females ($M = 1.85$, $SD = 2.46$) reporting a slightly higher number of traits than males ($M = 1.25$, $SD = 2.08$).

Individual Differences in Trust Patterns

We began by testing the random intercept model (null model) to estimate the intraclass correlation coefficient (ICC) to determine how much of the variability in MUs transferred across the 15 rounds resulted from between-person differences. The ICC was .253, indicating 25.3% of the variance in the amount invested across rounds can be explained by properties of the individual. This ICC value is consistent with our expectations based on prior exploration of trust behaviour during the TG (Korsgaard et al., 2018), and the knowledge that our experimental design possessed three distinct trust phases.

Level 1 Analyses

Following the procedure established by Bliese and colleagues (Bliese & Lang, 2016; Bliese & Ployhart, 2002), for each of our absolute and relative analyses, we first generated a linear-only baseline DGM (random intercept model) to determine the pattern of change in trust for participants as a whole using the TIME variable.

$$Trust_{ti} = \pi_{0i} + \pi_{1i}TIME_{ti} + \pi_{2i}DT_{ti} + \pi_{3i}DS_{ti} + \pi_{4i}RT_{ti} + \pi_{5i}RS_{ti} + \varepsilon_{ti}$$

This model consisted of an intercept (π_{0i}), error variance (ε_{ti}), and the change variables: linear change over time ($TIME_{ti}$), dissolution transition (DT_{ti}), dissolution slope (DS_{ti}), restoration transition (RT_{ti}), and restoration slope (RS_{ti}), as described in Table 1.

We then compared the random intercept model, which describes the overall trajectory averaged across all participants, with a model that allows the trajectory of each participant to vary. By allowing for random variance across each change coefficient, we can estimate whether there are between-person differences in the pattern of MUs transferred. Beginning with the random intercept model, each consecutive model allows an additional change coefficient to freely vary across participants. Each consecutive model was tested against the previous model using the log-likelihood test (for each model, see rows 1 to 6 of Table 3a-c below). Results indicate that a model accounting for random effects for all change coefficients was the best fitting model for both the absolute and relative models, except for the dissolution slope in the model looking at change relative to the formation phase (see Table 3c). However, this does not affect the interpretation of this model since its purpose in the present study is to examine trust changes during the restoration phase in comparison to trust levels and growth before the violation took place. In the next step, we tested for lag-1 autocorrelation (row 7 of Table 3 for each model) and heteroscedasticity (row 8 of Table 3 for each model). Log-likelihood ratio tests indicated a significantly better fit only when we accounted for autocorrelation. Models that controlled for both error structures simultaneously did not converge.

Table 3*Model Comparison Tests for Discontinuous Growth Models, Autocorrelations, and Heteroscedasticity*

Model	df	AIC	BIC	logLik	Test	L.Ratio
a) Absolute Model						
1. Random Intercept Model	8	33229.79	33279.08	-16606.89		
2. Random TIME	10	33206.92	33268.54	-16593.46	1 vs 2	26.87***
3. Random TIME & DT	13	33114.75	33194.85	-16544.37	2 vs 3	98.17***
4. Random TIME, DT & DS	17	33072.36	33177.11	-16519.18	3 vs 4	50.39***
5. Random TIME, DT, DS & RT	22	33057.37	33192.92	-16506.68	4 vs 5	24.99***
6. Random TIME, DT, DS, RT & RS	28	32991.98	33164.50	-16467.99	5 vs 6	77.39***
7. Autocorrelation Error Structure	29	32968.31	33147.00	-16455.15	6 vs 7	25.67***
8. Heteroscedasticity	29	32992.60	33171.29	-16467.30	6 vs 8	1.38
b) Relative Model (preceding phase as a baseline)						
1. Random Intercept Model	8	33229.79	33279.08	-16606.89		
2. Random TIME	10	33113.18	33174.80	-16546.59	1 vs 2	120.60***
3. Random TIME & DT	13	33065.67	33145.77	-16519.83	2 vs 3	53.51***
4. Random TIME, DT & DS	17	33033.18	33137.93	-16499.59	3 vs 4	40.49***
5. Random TIME, DT, DS & RT	22	33002.31	33137.87	-16479.16	4 vs 5	40.87***
6. Random TIME, DT, DS, RT & RS	28	32992.08	33164.61	-16468.04	5 vs 6	22.23***
7. Autocorrelation Error Structure	29	32968.45	33147.14	-16455.22	6 vs 7	25.63***
8. Heteroscedasticity	29	32992.61	33171.30	-16467.31	6 vs 8	1.47
c) Relative Model (formation phase as a baseline)						
1. Random Intercept Model	8	33229.79	33279.08	-16606.89		
2. Random TIME	10	33113.18	33174.80	-16546.59	1 vs 2	120.60***
3. Random TIME & DT	13	33103.59	33183.70	-16538.80	2 vs 3	15.59**
4. Random TIME, DT & DS	17	33109.18	33213.93	-16537.59	3 vs 4	2.42
5. Random TIME, DT, DS & RT	22	33037.07	33172.63	-16496.54	4 vs 5	82.11***
6. Random TIME, DT, DS, RT & RS	28	32992.06	33164.58	-16468.03	5 vs 6	57.01***
7. Autocorrelation Error Structure	29	32968.41	33147.10	-16455.21	6 vs 7	25.65***
8. Heteroscedasticity	29	32992.67	33171.35	-16467.33	6 vs 8	1.39

DT = dissolution transition, DS = dissolution slope, RT = restoration transition, RS = restoration slope.

* $p < .05$, ** $p < .01$, *** $p < .001$.

The Final DGM's in Table 4 below provides the parameter estimates and standard errors with random transitions and slopes, and autocorrelation included. The results of the absolute model indicate that, on average, the MUs transferred decreased significantly immediately following the trust violation ($est_{DT} = -20.60, p < .001$) and increased significantly immediately following the trust repair ($est_{RT} = 10.83, p < .001$). This confirms that the experimental manipulation of trustee reciprocity to create trust violation and trust restoration was successful. There was no evidence of a linear trend in MUs transferred during the formation ($est_{FS} = -.12, n.s.$) and dissolution phases ($est_{DS} = -.80, n.s.$), but there was a positive linear trend in MUs transferred during the restoration phase ($est_{RS} = 1.26, p < .01$).

The results of the corresponding relative model further indicate that the immediate shift in MUs transferred following the trust violation ($est_{DT} = -20.49, p < .001$) resulted in transfer amounts that were significantly lower than what would be expected if the violation had not occurred. Furthermore, the immediate shift in the MUs transferred following the trust repair ($est_{RT} = 11.63, p < .001$) resulted in transfer amounts that were significantly higher than what would be expected if the repair had not occurred. However, when using the formation phase as a baseline, this increase in transfer amounts at repair was still significantly lower than what would be expected if neither the violation nor repair had occurred ($est_{RT} = -10.91, p < .01$). Regarding trust growth, the slope for the dissolution phase was not found to differ significantly from the slope during the formation phase ($est_{DS} = -.69, n.s.$). However, the positive linear trend in MUs transferred during restoration was steeper than that observed during the formation phase ($est_{RS} = 1.38, p < .10$) and during the dissolution phase ($est_{RS} = 2.07, p < .10$).

Table 4*Final Discontinuous Growth Models*

Variables	Absolute		Relative (to preceding phase)		Relative (to formation phase)	
	Est	SE	Est	SE	Est	SE
Intercept	46.42***	1.51	46.42***	1.51	46.42***	1.51
Formation slope (FS)	-0.12	0.58	-0.12	0.58	-0.12	0.58
Dissolution transition (DT)	-20.60***	2.39	-20.49***	2.79	-20.49***	2.79
Dissolution slope (DS)	-0.80	1.16	-0.69	1.28	-0.69	1.28
Restoration transition (RT)	10.83***	1.87	11.63**	2.63	-10.91**	4.15
Restoration slope (RS)	1.26**	0.40	2.07 [†]	1.25	1.38 [†]	0.71

[†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$, tests are two-tailed, $n=234$ participants, 3510 observations.

The results of the absolute model indicate that, on average, the MUs transferred decreased significantly immediately following the trust violation ($est_{DT} = -20.60$, $p < .001$) and increased significantly immediately following the trust repair ($est_{RT} = 10.83$, $p < .001$). This confirms that the experimental manipulation of trustee reciprocity to create trust violation and trust restoration was successful. There was no evidence of a linear trend in MUs transferred during the formation ($est_{FS} = -.12$, *n.s.*) and dissolution phases ($est_{DS} = -.80$, *n.s.*), but there was a positive linear trend in MUs transferred during the restoration phase ($est_{RS} = 1.26$, $p < .01$). The results of the corresponding relative model further indicate that the immediate shift in MUs transferred following the trust violation ($est_{DT} = -20.49$, $p < .001$) resulted in transfer amounts that were significantly lower than what would be expected if the violation had not occurred. Furthermore, the immediate shift in the MUs transferred following the trust repair ($est_{RT} = 11.63$, $p < .001$) resulted in transfer amounts that were significantly higher than what would be expected if the repair had not occurred. However,

when using the formation phase as a baseline, this increase in transfer amounts at repair was still significantly lower than what would be expected if neither the violation nor repair had occurred ($est_{RT} = -10.91, p < .01$). Regarding trust growth, the slope for the dissolution phase was not found to differ significantly from the slope during the formation phase ($est_{DS} = -.69, n.s.$). However, the positive linear trend in MUs transferred during restoration was steeper than that observed during the formation phase ($est_{RS} = 1.38, p < .10$) and during the dissolution phase ($est_{RS} = 2.07, p < .10$)

Level 2 Analyses

To test for systematic differences in MUs transferred between individuals due to differences in the number of BPD traits reported, BPD trait count was added as a Level 2 predictor for each of the Level 1 components. The associated Level 2 equations are as follows:

$$\begin{aligned}\pi_{0i} &= \beta_{00} + \beta_{01}(BPD)_i + r_{0i} \\ &\vdots \\ \pi_{5i} &= \beta_{50} + \beta_{51}(BPD)_i + r_{5i}\end{aligned}$$

Based on results from separate main effects models for gender ($est_{Female} = -5.64, p < .01$) and cognitive reflective ability ($est_{Cognitive} = 2.52, p < .01$) which found that being female or having low cognitive reflective ability were associated with smaller transfers overall, we controlled for both in our final interaction models. Results for the final discontinuous mixed-effects models are presented in Table 5.

Table 5

Discontinuous Mixed-Effects Growth Models Predicting Trust as a Function of BPD Trait Count after Controlling for Gender and Cognitive Reflective Ability

	Absolute		Relative (preceding phase)		Relative (formation phase)	
	Est	SE	Est	SE	Est	SE
Intercept	49.14***	2.02	49.12***	2.02	49.20***	2.02
Formation slope (FS)	-0.08	0.57	-0.08	0.57	-0.08	0.57
Dissolution transition (DT)	-20.74***	2.40	-20.67***	2.79	-20.67***	2.80
Dissolution slope (DS)	-0.76	1.16	-0.69*	1.27	-0.69	1.27
Restoration transition (RT)	10.55***	1.86	11.31**	2.63	-11.42**	4.12
Restoration slope (RS)	1.29**	0.40	2.05	1.26	1.36	0.71
Gender (Female)	-4.30*	2.15	-4.28*	2.15	-4.41*	2.15
Cognitive reflective ability	2.34*	1.04	2.34*	1.04	2.36*	1.03
BPD	0.17	1.49	0.17	1.49	0.18	1.50
FS * BPD	-1.12 [†]	0.57	-1.12 [†]	0.57	-1.12 [†]	0.57
DT * BPD	2.13	2.40	3.25	2.79	3.25	2.79
DS * BPD	2.33*	1.16	3.45**	1.27	3.45*	1.27
RT * BPD	-3.83*	1.85	-6.16*	2.62	7.43 [†]	4.11
RS * BPD	0.43	0.40	-1.90	1.25	1.55*	0.71

BPD = Number of BPD traits reported on MSI-BPD. BPD and cognitive reflective ability were z -standardized and centered at the sample mean.

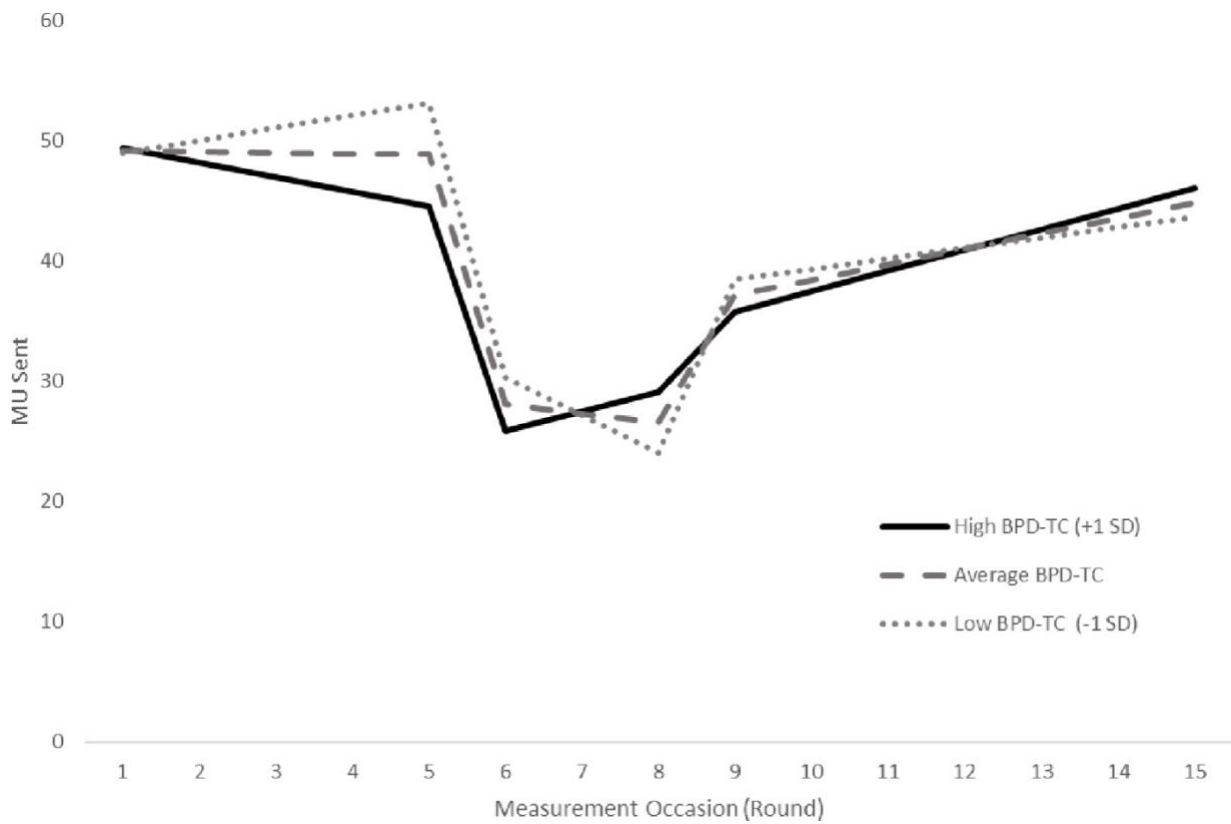
[†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$, tests are two-tailed, $n=234$ participants, 3510 observations.

Figure 2 graphs the effects of BPD on the overall change pattern of MUs transferred. MUs transferred at each measurement occasion was predicted for individuals with a high (1 SD above the sample mean) and low (1 SD below the sample mean) BPD trait count,

contrasted with predicted MUs transferred for individuals scoring at the sample mean of BPD. Results for the absolute model are reported first, followed by an outline of any notable deviations observed when looking at relative change.

Figure 2

Number of MUs Sent by the Trustor as a Function of Borderline Personality Disorder Trait Count



Note. High BPD trait count are trustors who scored one or more standard deviations above the mean number of BPD traits on the McLean Screening Instrument for Borderline Personality Disorder, and Low BPD trait count are trustors who scored one or more standard deviations below the mean number of BPD traits.

Results do not indicate a main effect of BPD trait count. With respect to changes in MUs transferred, it was hypothesized that BPD trait count would have a deleterious effect on the amount transferred at each transition, representing acute changes in trust in reaction to violation and repair, and on the rate of change of MUs transferred during each phase, representing the rate of trust growth or decline.

Formation Slope

We predicted that BPD would be associated with a slower rate of trust growth during the formation phase. As indicated by Table 5, this hypothesis was confirmed for the relationship between BPD trait count and the trajectory of trust during the formation phase where we found a significant negative linear trend ($est_{FS*BPD} = -1.12, p < .10$). This result indicates that individuals with a higher number of BPD traits have a more pronounced decline in MUs transferred during the trust formation phase. This was additionally confirmed in a model which isolated only the first 5 rounds of the game to measure trust change ($est_{FS*BPD} = -1.02, p < 0.10$).

Dissolution Transition

It was predicted that BPD would be associated with a larger immediate decrease in the level of trust immediately after the trust violation. The hypothesis for the relationship between BPD trait count and the dissolution transition was not supported. While there was a decrease in MUs transferred following the first exposure to the trust violation ($est_{DT} = -20.74, p < .001$), the size of the decrease appeared to be similar regardless of BPD trait count ($est_{DT*BPD} = 2.13, n.s.$).

Dissolution Slope

It was hypothesized that BPD would be associated with a slower rate of trust growth during the dissolution phase. As indicated by Table 5, while BPD trait count appeared to be a substantive moderator on the change in MUs transferred during the dissolution phase, the effect was not in the hypothesized direction. Rather, there was a significant positive linear trend between BPD trait count and the amount transferred during dissolution ($est_{DS*BPD} = 2.33, p < .05$). This result indicates that individuals with a higher number of BPD traits have a more pronounced increase in the rate of MUs transferred during the dissolution phase, rather than a decline as predicted.

The comparative relative model is used to understand the impact of the trust violation on the intra-individual patterns of trust. It was hypothesized that BPD would be associated with a slower rate of trust growth during the dissolution phase relative to the trust growth observed in the formation phase. Once again, BPD was significantly associated with the rate at which trust changed during dissolution relative to formation, but not in the hypothesized direction ($est_{DS*BPD} = 3.45, p < .01$). In relative terms, BPD was associated with a rate of growth in MUs transferred during dissolution that was greater than the rate of growth during formation.

Restoration Transition

It was predicted that BPD would be associated with a smaller increase in the level of trust immediately following the trust repair. As hypothesized, BPD trait count appeared to be a substantive moderator on the increase in the number of MUs transferred at the restoration transition ($est_{RT*BPD} = -3.83, p < .05$). High BPD trait individuals had a significantly less pronounced increase in MUs transferred following the first instance of repair. This effect was

even more pronounced when considering the shift in trust levels following repair relative to the level of trust predicted had the repair never occurred ($est_{RT*BPD} = -6.16, p < .05$). In other words, for persons high on BPD trait count, trust levels increased immediately following restoration but to a significantly less extent than persons low on BPD trait count. Finally, when considering the shift in trust levels following repair relative to the formation phase, BPD was associated with a larger number of units transferred ($est_{RT*BPD} = 7.43, p < .10$). In other words, the level of trust (at round 9) was greater than what would have been expected based on the formation phase pattern of trust change.

Restoration Slope

It was predicted that BPD would be associated with a slower rate of trust growth during the restoration phase. Our hypothesis for the relationship between BPD trait count and the restoration slope was not supported. The linear pattern in MUs transferred during the restoration phase ($est_{RS} = 1.29, p < .01$) was at a similar rate regardless of BPD trait count ($est_{RS*BPD} = .43, n.s.$).

With reference to relative change, it was predicted that BPD would be associated with a less pronounced rate of trust growth during the restoration phase relative to the formation phase. However, while BPD trait count was found to significantly moderate the rate of relative change during restoration, it was not in the hypothesized direction. BPD trait count was associated with a more pronounced positive linear trend in the number of MUs transferred during restoration relative to the linear trend in transfers observed in the formation phase ($est_{RS*BPD} = 1.55, p < .05$). BPD was associated with a rate of growth in MUs transferred during restoration that was greater than the rate of growth during formation.

It was also predicted that BPD would be associated with a less pronounced rate of trust growth during the restoration phase even after taking into account the trajectory during dissolution. Results indicate a non-significant coefficient trending in the hypothesized direction ($est_{RS*BPD} = -1.90, p = .13$). It should be noted however that we expected BPD trait count to be associated with faster dissolving trust in response to the violation, but we found the opposite; indeed, trust grew at a faster rate. In light of this, the non-significant trend described above suggests that, in comparison to the faster growth observed in the dissolution phase, trust growth was tempered during restoration at higher BPD trait counts.

Trust change as a function of BPD caseness

To support the clinical utility of our findings the level 2 analyses were repeated using the MSI-BPD recommended cutoff values to group participants into likelihood of “caseness” ($N = 16$) and “noncaseness” ($N = 218$). Results reported are limited to key findings, but a complete account can be found in Table 6. In contrast to the BPD trait model, BPD caseness was not found to modify the rate of trust growth during the formation phase, although there were non-significant trends in the expected direction ($est_{FS*caseness} = -3.57, p = .12$). BPD caseness was found to positively moderate trust growth during the dissolution phase, both relative to the formation phase ($est_{DS*caseness} = 14.32, p < .01$) and in absolute terms ($est_{DS*caseness} = 10.74, p < .05$).

Table 6*Discontinuous Mixed-Effects Growth Models Predicting Trust as a Function of BPD Caseness after Controlling for Gender and Cognitive Reflective Ability*

	Absolute		Relative (to preceding phase)		Relative (to formation phase)	
	Est	SE	Est	SE	Est	SE
Intercept	48.97***	2.05	48.98***	2.06	49.00***	2.06
Formation slope (FS)	0.17	0.60	0.17	0.60	0.17	0.60
Dissolution transition (DT)	-20.71***	2.49	-20.88***	2.90	-20.88***	2.90
Dissolution slope (DS)	-1.50	1.20	-1.67	1.31	-1.67	1.32
Restoration transition (RT)	11.51***	1.93	13.01***	2.71	-12.87**	4.28
Restoration slope (RS)	1.12**	0.41	2.61*	1.30	0.95	0.73
Gender (Female)	-4.48*	2.14	-4.49*	2.15	-4.53*	2.15
Cognitive reflective ability	2.41*	1.04	2.40*	1.04	2.40*	1.04
BPD caseness	4.09	5.88	4.08	5.92	4.08	5.91
FS * BPD caseness	-3.57	2.27	-3.57	2.28	-3.57	2.27
DT * BPD caseness	-0.39	9.52	3.19	11.07	3.19	11.06
DS * BPD caseness	10.74*	4.59	14.32**	5.02	14.31**	5.03
RT * BPD caseness	-14.08†	7.34	-24.82*	10.34	21.32	16.34
RS * BPD caseness	2.45	1.57	-8.29†	4.96	6.02*	2.80

† $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$, tests are two-tailed, $N=234$ participants ($n=16$ 'caseness', $n=218$ 'noncaseness'), 3510 observations.

BPD caseness = 7 or more BPD traits reported on MSI-BPD. Cognitive reflective ability was z -standardized and centred at the sample mean.

In response to the repair, BPD caseness was associated with a smaller absolute increase in MUs transferred ($est_{RT*caseness} = -14.08, p < .10$). For those meeting BPD caseness, trust levels were also lower after the first instance of repair than that predicted by the trajectory of growth during the dissolution phase ($est_{RT*caseness} = -8.29, p < .10$). However, caseness was not found to modify the level of trust following repair relative to the level expected based on the formation trajectory ($est_{RT*caseness} = 21.32, p = .19$). Finally, the rate of trust growth during restoration for those meeting BPD caseness was faster than trust growth during formation ($est_{RS*caseness} = 6.02, p < .05$), and slower than trust growth during dissolution ($est_{RS*caseness} = -8.29, p < .01$).

Discussion

The present study used DGM with a TG to investigate the relationship between BPD and trust behaviours. It examined how BPD trait count modified the level and trajectory of trust as it formed with a new partner, dissolved in response to trust violation, and was rebuilt in response to trust repair. The results suggest that trust behaviour in individuals with a high number of BPD traits may be broadly classified as cautious and mistrustful in the beginning of a new relationship, even when the other party is behaving in a cooperative and trustworthy manner, and conversely, during and after a trust violation, trust appears to grow markedly, even in the face of repeated betrayals.

When interacting with a new and cooperative partner, high BPD trait individuals became progressively less trusting. That is, despite earning up to 32% profit on each investment made during this period, as the number of BPD traits increased, individuals progressively reduced the number of MUs transferred. A trend in this direction was also

observed in those whose trait levels indicated likely caseness for BPD. This result augments previous findings indicating that average trust levels and trust growth are lower for people with BPD when playing the TG with a new and not uncooperative partner (Liebke et al., 2018; Unoka et al., 2009).

Regarding the impact of BPD traits on trust behaviour following violation and repair, the findings highlight the benefits of using a method of analysis that is responsive to the dynamic, multiphasic nature of trust. While previous research found average trust levels during dissolution and restoration to be comparable between BPD patients and HCs (Liebke et al., 2018), the current study found BPD had a paradoxical influence on trust patterns after violation and repair. If increasing mistrust defines the pre-violation trust patterns of high BPD trait count individuals, incessant trust growth would best characterise the violation and post-violation periods.

First, despite experiencing three consecutive rounds in which the trustee kept all or most of the investment, BPD trait count was associated with making progressively larger transfers, a finding contrary to what we expected based on King-Casas and colleagues observation that trustees with BPD were less likely to respond to diminishing cooperation via increasing reciprocity (King-Casas et al., 2008). Given that the immediate reduction in the amount transferred following the first instance of violation was of a comparable magnitude across all levels of BPD symptomatology, even when the pre-violation trajectories were accounted for, it is likely that the three violation rounds were recognized as a norm violation irrespective of level of borderline pathology. This is consistent with findings that individuals with BPD can accurately appraise the fairness of trustee reciprocity (Franzen et al., 2011). It is possible that the decreases in investment size signalling diminishing trust were more

nuanced in the King-Casas study (2008), whereas our trust violation was unambiguous and therefore, may have prompted an unconventional response. The disparate coaxing behaviours between King-Casas' trustees and our trustors may also reflect a property of the roles. It is possible that the powerlessness inherent in the trustee role – it is the trustor who decides whether to initially engage – elicits a more aggressive response in people with BPD, while as trustor, they can choose whether to be benevolent knowing that they can stop engaging at any point.

Regarding trust behaviours following repair, the size of the increase in MUs transferred following the initiation of trust repair became less pronounced as trait count increased. For persons who reported 9 or 10 BPD traits, there was a decrease rather than the expected increase. In other words, those with the highest levels of BPD traits paradoxically reduced the size of their next transfer in response to the trustee's return to cooperative play. It is possible that the first sign of reparative action by the trustee elicited caution and suspicion in higher trait individuals. In fact, if a prediction were to be made about how many MUs high BPD individuals would have sent at that point, had the violation continued, it would be a markedly higher amount than that sent in response to the repair. On the other hand, perhaps individuals with a higher BPD trait count had sought to coax higher returns from the trustee, and upon achieving this objective, stopped the coaxing behaviour and reduced their investments. Finally, it was found that trustors, in general, were on average observed to invest progressively larger amounts throughout the restoration phase. When this growth was considered relative to the growth observed in the formation phase, BPD traits were associated with a comparatively faster rate of growth. That is, trust in high BPD trait persons was observed to restore at a more accelerated rate than it was formed in the period before the

violation took place. In fact, as trait count increased, trust grew at a faster rate during both dissolution and restoration than it had during formation.

This pattern of intraindividual fluctuations does not appear to map on to the intraindividual trust fluctuations of the general population (Fulmer & Gelfand, 2015; Korsgaard et al., 2018), with trust during each phase largely appearing to flow in the opposite direction. This suggests that although the ability to recognize norm violations does not appear to be compromised, BPD is associated with intraindividual changes in *trust behaviours* that are socially atypical. This has serious implications for individuals with BPD in terms of how they may be experienced by others during interpersonal interactions. Social cognition processes are believed to be engaged when individuals make strategic interpersonal decisions (Frith & Singer, 2008; Hampton et al., 2008; Lee & Harris, 2013). To maintain a mutually beneficial equilibrium in the TG, each party must be able to recognize, decipher, and respond appropriately to the signals sent. In the case of repeated interactions with the same partner such as in an iterated TG, the intraindividual fluctuations in behaviour communicate meaningful interpersonally relevant data.

In addition to being able to model the mind of the other, Kishida and colleague (2010) also propose that fair social exchange requires three computational capacities in each agent: to compute the social norms associated with such an exchange, to recognize deviations from said norms, and to respond appropriately considering these deviations. In other words, each agent must not only recognize and ascribe meaning to the intraindividual fluctuations of social signals emitted by their partner, but also determine if these fluctuations are socially normative. For example, previous research using a college student population found that the normative pattern in intraindividual fluctuations is that the *restoring* of trust appears to be a

lengthier process than forming trust with a new partner (Korsgaard et al., 2018). The socially non-normative pattern of intraindividual fluctuations associated with BPD symptomatology may be perceived by others as incomprehensible, unexpected, and perhaps even odd interpersonal dynamics. In as few as 15 rounds of social exchange, the current study showed that borderline pathology predicts a paradoxical relational style that may invite confusion, whereby betrayal begets trust and cooperation, mistrust.

It is conceivable that these kinds of behaviours and preferences are likely to compromise the development of healthy relationships and may lead to relationship breakdowns or attract partners who may perpetuate these potentially deleterious relational dynamics (Bouchard et al., 2009). Our results also support previous research in which individuals with BPD have been found to demonstrate greater acceptance of and perhaps a preference for inequitable treatment. For instance, findings from economic game studies suggest that compared to HCs, BPD individuals are more likely to accept unfair offers (Polgar et al., 2014) and reject fair offers (De Panfilis et al., 2019), and express a greater preference for an unfair interaction partner, and lower preference for a fair interaction partner (Jeung et al., 2020). While much of the trust literature in borderline populations has focused on BPD tendencies towards mistrust and lack of cooperation, our analytical methodology has highlighted the other extreme, the tendency to engage in trusting behaviour in contexts warranting prudence. Indeed, findings from a social network analysis study has showed that BPD is associated with reduced discrimination in differentiating whom in their social network they seek advice and emotional support from (Clifton et al., 2007). That is, despite the tendency towards trust-compromising beliefs, appraisals, and behaviours, people with BPD may also trust more haphazardly or arbitrarily.

Limitations & future directions

The current study has several limitations. First, a clinical sample was not used. Previous research has primarily used patients with a BPD diagnosis, so it is possible our findings may not apply to individuals with severe and persistent BPD impairments. It is important to note however, that interpersonal disturbances in nonclinical samples of people with borderline traits are almost as profound as in clinical samples (Tolpin et al., 2004), and that young adults with sub-clinical borderline features are more likely to exhibit interpersonal dysfunction at a two-year follow up than their healthy counterparts (Trull et al., 1997).

Second, we used a simulated trustee rather than an actual human being, and therefore created interactions where the trustee was unconditionally cooperative or uncooperative, regardless of our participants' behaviour. While standardized trustee reciprocity was chosen intentionally to create the distinct phases of trust, doing so limits the conclusions that can be drawn since interpersonal trust is a dyadic process (Simpson, 2007). Third, participants were not offered financial incentives to participate. By not tying the game results to a financial reward, participants may have been less motivated to take the game seriously. Fourth, BPD is most typically associated with relational disturbances in close relationships. It is not clear whether the behaviour observed in a low stakes game with an anonymous partner would reflect trust behaviours in close personal relationships. For example, Miano et al. (2017) found that within the context of intimate relationships, individuals with BPD compared to HCs appraised their romantic partner as less trustworthy after discussing a relationship threatening or personally threatening topic, whereas appraisal ratings were comparable following discussion of a neutral topic. It is therefore likely that the experience of a trust violation and repair within the context of an intimate relationship may evoke a more marked or varied response.

Our findings and methodology provide a solid foundation upon which researchers can examine the nuanced factors and processes that may underlie these incongruous trust dynamics, and therefore help inform more targeted interventions. For example, previous research has looked at the effect of underlying attachment insecurity on trust behaviours (e.g., Bartz, Simeon, et al., 2011; M. J. McClure et al., 2013; Taheri et al., 2018). Our methodology could be used to elicit potentially attachment salient events such as a trust violation and examine whether attachment insecurity underpins the effect of borderline pathology on trust behaviours. Our findings should also be replicated using a patient sample to allow clinical inferences to be made.

Conclusions

To the best of our knowledge, this is the first study with a BPD focus to use DGM with a TG to examine trust as a dynamic and multiphasic process. Specifically, the study revealed the trajectories of trust as it forms, dissolves, and restores in response to trust violation and repair. It also explained how these trajectories varied as a function of the number of BPD traits reported. Showing how trust fluctuates within the individual and how trait count modifies the magnitude and direction of the fluxes, provides a richer and more nuanced understanding of how people with BPD traits engage with others in trust-altering contexts. This approach uncovered a paradoxical style of social exchange where social norms appear to be contradicted, thereby creating interpersonal encounters that are seemingly ambivalent, aberrant, and puzzling. By adopting a design and analytical methodology that recognizes the dynamic nature of trust, the study uniquely illustrated how relational disturbances may be produced and maintained in a BPD population.

Acknowledgments

The authors would like to thank Professors Paul Bliese and Jonas Lang for statistical consultation, and Yossef Abramov for developing the online trust game platform.

Chapter 3: The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count

This chapter has been published in *Journal of Psychopathology and Behavioral Assessment*. The chapter is identical to the published manuscript except for figure numbers (Figures 1 & 2 renamed as Figures 3 & 4, respectively), table numbers (Tables 1, 2, & 3 renamed as Tables 7,8, & 9, respectively), and replacing the term “symptom count” with “trait count”, which have been altered to ensure uniformity in formatting across the thesis. A number of other minor formatting changes have been made in the service of overall uniformity and readability.

Abramov, G., Kautz, J., Miellet, S., & Deane, F. P. (2022). The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count. *Journal of Psychopathology and Behavioral Assessment*. <https://doi.org/10.1007/s10862-022-09965-9>

Abstract

Borderline personality disorder (BPD) is associated with paradoxical trust behaviours, specifically a faster rate of trust growth in the face of trust violations. The current study set out to understand whether attachment style, self-protective beliefs, and feelings of rejection underpin this pattern. Young adults ($N=234$) played a 15-round trust game in which partner cooperation was varied to create three phases of trust: formation, dissolution, and restoration. Discontinuous growth modelling was employed to observe whether the effect of BPD trait count on trust levels and growth is moderated by fearful or preoccupied attachment style, self-protective beliefs, and feelings of rejection. Results suggest that the slower rate of trust formation associated with BPD trait count was accounted for by feelings of rejection or self-protective beliefs, both of which predicted a slower rate of trust growth. The faster rate of trust growth in response to trust violations associated with BPD trait count was no longer significant after self-protective beliefs were accounted for. Interventions targeting self-protective beliefs and feelings of rejection may address the trust-based interpersonal difficulties associated with BPD.

Introduction

Relational disturbances are a hallmark feature of borderline personality disorder (BPD) and are present among subclinical populations (Tolpin et al., 2004), persisting even after symptoms remit (Zanarini et al., 2010). Individuals with BPD tend to have intense relationships marred by conflict, instability, and rupture (Bouchard et al., 2009; Clifton et al., 2007; Lazarus et al., 2020). Aberrant trust processes have been proposed as a contributor to impaired interpersonal functioning in BPD (Poggi et al., 2019). Trust behaviours have been examined using behavioural economic experimental paradigms such as the trust game (TG: Berg et al., 1995). Evidence suggests that in these TG interpersonal exchanges, BPD is associated with reduced cooperation (King-Casas et al., 2008) and increased mistrust (Unoka et al., 2009), even when engaging with a cooperative partner (Abramov et al., 2020; Liebke et al., 2018).

Currently, very little is known about the mechanisms that underpin the trust practices of individuals with BPD. Individuals with BPD typically develop mistrustful models of others, endorsing beliefs that others will betray, exploit, and deceive (Barazandeh et al., 2016; Bhar et al., 2008). As such trust behaviours among individuals with BPD may be modulated by these beliefs. There is also growing evidence that situations or states which typically promote prosocial behaviours appear to have a paradoxical effect among individuals with borderline pathology (Abramov et al., 2020; Ebert et al., 2013; Liebke et al., 2018). For example, individuals with BPD appear to engage in less trusting behaviours following social acceptance than following social rejection (Liebke et al., 2018), suggesting rejection experiences may play a mechanistic role in trust processes in BPD. More recently, evidence suggests individuals who endorse many BPD symptoms appear to withhold trust when engaging with a partner whose actions appear trustworthy and cooperative, yet make more

trusting gestures when that partner becomes uncooperative (Abramov et al., 2020). A similarly paradoxical effect was observed among individuals with BPD who compared to controls, appeared to become less trusting following administration of the neuropeptide oxytocin, known for its prosocial effects on behaviour and used by researchers to activate the attachment system (Ebert et al., 2013), highlighting the potential role of attachment disturbance in modulating trust behaviours among individuals with BPD. The current study aims to understand the mechanisms underpinning the paradoxical trust behaviours associated with BPD by examining whether attachment style, self-protective beliefs, and feelings of rejection moderate the effects of BPD trait count on trust behaviours in response to cooperative and uncooperative partner play in a social exchange.

A brief overview of the TG will be provided followed by a detailed rationale for the relationship between the study variables, BPD, and trust. In the TG, trust is operationalized as the proportion of a monetary endowment (monetary units, MU) an individual, known as the *trustor*, is willing to entrust to an anonymous party, the *trustee* (Berg et al., 1995). The amount sent is multiplied by a factor before being received by the trustee who then has the option of returning a proportion to the trustor, which may result in a loss or profit to the trustor. Previous research has used the TG to examine trust as a dynamic phenomenon whereby trustee reciprocity (trustworthiness) has been modified to create phases of trust (e.g., Abramov et al., 2020; Audrey Korsgaard et al., 2018; Fulmer & Gelfand, 2013; Liebke et al., 2018). Specifically, these are *trust formation* (initial stage of TG where trustee reciprocity is held at levels that incur no loss and potential gain to the trustor), *trust dissolution* (trustee reciprocity results in loss to the trustor, constituting a violation of trust), and *trust restoration* (trustee reciprocity returns to pre-violation levels, constituting gestures of trust reparation).

Attachment style and trust

The socially atypical vicissitudes in trust associated with borderline pathology can be understood through the lens of attachment theory (Bowlby, 1969, 1973, 1980). Aetiological models of BPD suggest it is a disorder that in part stems from attachment disturbances arising from suboptimal, adverse, or invalidating caregiver experiences (Fonagy et al., 1995; Gunderson & Lyons-Ruth, 2008; Linehan, 1993). Interpersonal trust involves the belief that a person cares about one's needs and can be depended upon (Rempel et al., 1985) and securely attached individuals believe that significant others will be available, caring, and responsive, leading Mikulincer (1998) to conclude that trust was an integral tenet of secure attachment. The adult attachment styles associated with BPD are fearful-avoidant attachment, and to a lesser extent, anxious-preoccupied attachment (Agrawal et al., 2004). Both styles endorse a negative model of the self in terms of personal insecurity, but those endorsing a more preoccupied style seek intimacy and connection while those endorsing a fearful attachment style avoid or find it difficult to become close to and dependent on others (Bartholomew & Horowitz, 1991). Research looking at trust and attachment style has used a dual-dimensional conceptualization of adult attachment. The dimension of attachment-related anxiety, refers to the degree to which an individual is worried about being rejected by the other, and attachment-related avoidance, refers to the degree to which strategies are used to downregulate attachment needs in relational contexts, with high scores indicating a discomfort with being close or dependent on the other (Fraley et al., 2015). Both fearful and preoccupied attachment styles reflect increased attachment-related anxiety, but only the fearful attachment style also has elevated levels of attachment-related avoidance.

The findings on the effect of attachment insecurity on trust behaviours are varied, and almost exclusively studied in non-clinical populations. Anxiously attached individuals tend

be more hesitant and more mistrustfully inconsistent in their responses in a social dilemma game (M. J. McClure et al., 2013). Likewise, under conditions of uncertainty where participants could randomly gain or lose money, participants became more cooperative, but this effect was muted among individuals with high levels of attachment anxiety or avoidance (Taheri et al., 2018). In contrast, there is some evidence that trust formation is positively associated with attachment anxiety (Fett et al., 2016). Furthermore, individuals with high levels of attachment anxiety may be more willing to self-sacrifice in favour of others. For example, in a bargaining game, anxiously attached individuals were both more generous in how much they were willing to offer the other player and more willing to accept less generous offers by the other player, in comparison to individuals with an avoidant attachment, who offered less and trended towards rejecting more offers (Almakias & Weiss, 2012). The authors suggested that anxiously attached persons are so concerned with gaining acceptance and avoiding abandonment, that they will forego monetary gain. On the other hand, they reasoned that avoidant individuals seek to avoid situations which might activate their attachment system, such as being exploited or unfairly rejected by the other. By making small offers, avoidant individuals can reason that any rejection that ensues was due to the low offer rather than a personal rejection (Almakias & Weiss, 2012). Curiously, this pattern observed in avoidantly attached individuals was found to reverse after administration of the neuropeptide oxytocin (De Dreu, 2012). However, the effects of oxytocin on those with a fearful attachment style appear to be detrimental to cooperation. In a mixed sample of patients with BPD and healthy controls (HC) who played a game that incentivised mutual cooperation, those with a preoccupied attachment style became more cooperative following administration of oxytocin, while individuals with a fearful attachment style became less cooperative (Bartz, Simeon, et al., 2011). This finding was replicated by Ebert (2013) who found a similar effect among individuals with BPD. This suggests that a fearful attachment

style may promote distrust and reduced willingness to cooperate under conditions of attachment arousal.

The current study examines whether attachment style underpins trust behaviours associated with BPD, and if so, whether there is a different effect based on the nature of the attachment disturbance. The extant research suggests that attachment anxiety is associated with heightened ambivalence about trusting but greater willingness to self-sacrifice in interpersonal transactions. Attachment avoidance appears to be associated with increased mistrust and a greater likelihood of behaving in a self-protective manner. Activation of attachment arousal also appears to have an adverse effect on trust behaviours for fearfully attached individuals, who report high levels of both attachment anxiety and avoidance. Endorsement of a fearful attachment style is expected to amplify the negative relationship between BPD and trust, thereby promoting lower levels of trust, while endorsement of a preoccupied attachment style is expected to mitigate that relationship, promoting higher levels of trust.

Self-protective beliefs and trust

Individuals with BPD endorse the belief that others will deceive, betray, and exploit (Barazandeh et al., 2016; Bhar et al., 2008). While none of the studies in the BPD literature have specifically examined the effect of holding such a priori beliefs on behavioural trust, there is evidence that in the absence of feedback regarding actual partner trustworthiness during a TG, BPD is associated with greater mistrust (Unoka et al., 2009). Dubbing these findings as representative of ‘unbiased trust,’ Unoka and colleagues (2009) also found that patients with BPD predict a less favourable outcome of the game than controls, suggesting that their comparatively mistrustful behaviour and predictions may reflect a disposition

towards perceiving others as untrustworthy. The current study will examine the effect of the apriori belief that one must pre-emptively act to protect oneself from betrayal by others on the relationship between BPD and trust behaviours. It is anticipated that self-protective beliefs will amplify the negative relationship between BPD and trust behaviours.

Feelings of rejection and trust

Interpersonal trust may also be influenced by negative affect that is present before the interpersonal exchange or elicited during the exchange. Events related to social rejection or abandonment in particular are known precipitants of emotional distress in BPD (Chapman et al., 2015; Staebler et al., 2011; Stiglmayr et al., 2005). Increased negative emotional arousal has been demonstrated to modify social processing in BPD (Dziobek et al., 2011; Wolff et al., 2007). Poggi and colleagues (2019) proposed that the mistrustful appraisals that are believed to underpin relational disturbances in BPD, may do so in conjunction with rejection sensitivity, another known mechanism of atypical social processing in BPD populations (Foxhall et al., 2019). Rejection sensitivity refers to a cognitive-affective processing disposition in which inevitable rejection by others is anxiously presumed, readily perceived, and overreacted to, often with hostility, attempts to control the other, or withdrawal (Downey & Feldman, 1996). Individuals high in rejection sensitivity expect that others will reject them and approach relationships with hypervigilance and hypersensitivity to signs of potential rejection, responding to actual or perceived rejection in ways that may compromise the relationship.

Rejection sensitivity has been found to mediate the relationship between BPD traits and an untrustworthiness bias for appraisals of neutral unfamiliar faces (Miano et al., 2013; Richetin et al., 2018). However, when it comes to trust behaviour, rejection and acceptance

appear to have a paradoxical effect for individuals with BPD. For example, Liebke et al. (2018) primed participants with a social activity in which they experienced either social acceptance or social rejection. Among participants with BPD, those who had been primed with acceptance feedback invested significantly less than those primed with feedback of rejection. The more positive the feedback was, relative to what was expected, the smaller the corresponding investment. The authors suggested that receiving feedback of social acceptance, particularly when one expected to be rejected, triggered defences in those with BPD resulting in more withholding behaviour (Liebke et al., 2018). The current study will examine the impact pre-existing feelings of rejection on the relationship between BPD and trust behaviours. It is predicted that baseline feelings of rejection moderate the relationship between BPD and trust behaviours by increasing mistrustful behaviours during the cooperative phases of the TG and increasing trustful behaviours during the uncooperative phases of the TG.

Finally, there is evidence suggesting that females exhibit less trusting behaviours in the TG (Johnson & Mislin, 2011). Given the gender bias in prevalence of BPD (American Psychiatric Association, 2013), the main effects of gender will be controlled. Cognitive reflective ability, that is, the ability to engage in conscious deliberation rather than respond impulsively (Frederick, 2005), has also been associated with a greater propensity to trust in the TG (Corgnet et al., 2016). As BPD has been associated with deficits in executive functioning impairments (see G. McClure et al., 2016; Ruocco, 2005) and social problem-solving (see Lazarus et al., 2014), altered decision making (see Paret et al., 2017), and increased impulsivity in interpersonal contexts (Berenson et al., 2016), the main effects of cognitive reflective ability will be controlled.

Method

The study was approved by the University of Wollongong ethics committee (HE2017/253). All participants provided informed consent. These data are part of a larger research project. One publication (Abramov et al., 2020) was based on an earlier exploration of this data, with the focus of that study being to build the preliminary model of BPD influence on trust growth. The current study is focused on data not previously published.

Participants

Participants ($N = 234$; 64% female; $M = 20.87$ years, $SD = 5.66$ years) were undergraduate students from a large Australian university who were invited to take part in an online study looking at the relationship between economic decision making and personality variables in exchange for course credit.

The trust game

The current study used a 15-round version of the TG (Berg et al., 1995) with trust operationalized as the proportion of allocated monetary units (MU) sent to a trustee for investment (Ben-Ner & Halldorsson, 2010). At the start of each of round the trustor was allocated 100 MU by the experimenter and given the option to send the trustee any proportion from 0-100% for investment. The amount sent was automatically tripled, and the trustee could repay any proportion from 0-100% of the tripled investment to the trustor.

Procedure

Participants were all assigned the role of trustor and were paired with another anonymous participant who was, in fact, a computer program. Trustee repayments were programmed so that following rounds 1-4 and 8-14, repayments were randomized to fall

between 34-44% of the tripled investment, providing the trustor with a return the equivalent of the original investment plus up to 32% profit². Following rounds 5-7, trustee repayments were randomized to fall between 0-10% of the tripled investment, providing the trustor with a return equivalent to losing 70% to 100% of their original investment³. This loss was designed to simulate a trust violation. Based on this repayment schedule, trustor investments can be divided into three distinct trust phases: formation (rounds 1-5), dissolution (rounds 6-8), and restoration (rounds 9-15). Overall, the average MU invested per round for each of the trust phases was 46.16 (SD = 18.95) during formation, 24.54 (SD = 19.14) during dissolution, and 38.4 (SD = 23.29) during restoration.

Measures

Trust

The number of MU's (0-100) that participants sent to the trustee in each round, represents a single behavioural measure of trust. Each participant provided 15 trust measures in total.

BPD Trait Count

The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD; Zanarini et al., 2003) was used to assess DSM-5 BPD symptomatology. The MSI-BPD is a 10-item screening instrument for BPD, with very good internal consistency (Cronbach's $\alpha = .84$, $N = 234$). BPD in the current study is operationalized as the number of MSI-BPD items endorsed (0-10), rather than a clinical diagnosis. The current sample endorsed a median of one symptom ($M = 1.63$, $SD = 2.34$). While MSI-BPD requires endorsement of 7 out of 10

² Profit ranged from 0 to 32% of investment.

³ Due to rounding to the nearest whole number, investments of \$1 did not incur a loss.

items for clinical diagnoses of BPD, endorsement of three or more items is reflective of sub-clinical borderline pathology (CLPS: Gunderson et al., 2011).

Attachment style

The Relationship Questionnaire (RQ; Bartholomew & Horowitz, 1991) is a 4-item questionnaire designed to measure adult attachment style. The RQ consists of four paragraphs, each describing an attitude toward relationships in representative of one of four attachment styles (secure, dismissing, preoccupied, and fearful). For the current study only the preoccupied and fearful attachment styles were reported. Preoccupied attachment was characterized as: “I am uncomfortable getting close to others. I want emotionally close relationships, but I find it difficult to trust others completely, or to depend on them. I worry that I will be hurt if I allow myself to become too close to others.” Fearful attachment was characterized as: “I am very comfortable without close emotional relationships. It is very important to me to feel independent and self-sufficient, and I prefer not to depend on others or have others depend on me.” Participants were asked to rate each paragraph on an 11-point Likert-type scale of 0 (not at all like me) to 10 (very much like me).

Self-protective beliefs

Self-protective beliefs were assessed using the Personality Beliefs Questionnaire—Borderline Personality Subscale (PBQ-BPD; Butler et al., 2002), a 14-item subset of the Personality Beliefs Questionnaire (PBQ; Beck & Beck, 1991). The PBQ-BPD subscale was developed on the basis of PBQ items that discriminated 84 BPD patients from 204 patients with other personality disorders (Bhar et al., 2008; Butler et al., 2002). Participants were asked to endorse each of the 14 items on a 5-point Likert-type scale from 0 (“I don’t believe it at all”) to 4 (“I believe it totally”). The subset has demonstrated adequate

internal consistency and discriminant validity (Butler et al., 2002). A study examining the factor structure of the PBQ-BPD using exploratory factor analysis found three factors: dependency, distrust, and the belief that one should take preemptive action to avoid threat (Bhar et al., 2008). In the current study, the three items comprising this preemptive action factor were used to provide a measure of self-protective beliefs (e.g., “People will get me if I don’t get them first”). Cronbach’s alpha was $\alpha = .74$, $N = 234$.

Feelings of rejection

Feelings of rejection were assessed using four items used previously in a study that assessed affective states in patients with BPD (Gadassi et al., 2014). Before playing the game, participants were prompted to rate on a 5-point Likert-type scale (0 = not at all, 4 = extremely) the extent to which they were currently experiencing specific emotions related to rejection (e.g., “At the present moment I feel rejected by others”). In the current study, internal consistency for pre-game rejection (Cronbach’s $\alpha = .74$, $N = 234$) was adequate, but lower than the .91 found in a sample of BPD patients (Gadassi et al., 2014).

Cognitive reflective ability

Cognitive reflective ability was measured using the Cognitive Reflection Test (CRT; Frederick, 2005), a three-item measure of the willingness to engage in deliberation during a cognitive task. Each item is a deceptively simple mathematical problem in which an intuitive/impulsive, yet incorrect answer must be suppressed in order to calculate the correct response. The number of items answered correctly were summed to provide a CRT score ranging from 0 – 3. The overall sample mean of 1.03 was consistent with previous findings (see Frederick, 2005).

Data Analyses

The NLME package (Pinheiro et al., 2019) included in the open source software R (R Development Core Team, 2018) was used to conduct mixed-effect discontinuous growth modelling (DGM) analyses (Bliese & Lang, 2016; Singer & Willett, 2003). These analyses assessed investment occasions (rounds) at Level 1 nested within individuals at Level 2. Level 1 parameters were coded based on the framework recommended by Singer and Willett (2003) and Bliese and Lang (2016) to create a matrix of time covariates that examine change in the average level of trust between each of the trust phases (i.e., formation, dissolution, and restoration) along with the growth of trust within each phase. To examine how individuals respond first to a trust violation and then to a trust restoration, change variables were such that the transition coefficients reflected the previous stage as the baseline for interpretation while the growth coefficients reflected growth relative to nil growth. *Time* was coded to capture the linear growth in trust during the formation phase. *Dissolution transition* (DT) and *Dissolution slope* (DS) were coded to represent the change in the average level of trust moving from the formation phase to the dissolution phase and the rate of trust growth during the dissolution phase, respectively. *Restoration transition* (RT) and *Restoration slope* (RS) were coded to represent the change in the average level of trust moving from the dissolution phase to the restoration phase and rate of trust growth during the restoration phase, respectively. The full DGM time-covariate matrix is presented in Table 7. Due to space constraints, the authors wish to encourage those interested in further understanding of the covariate matrix to read Bliese and Lang (2016) and/or Bliese, Kautz, and Lang (2020).

Table 7*Coding and Interpretation of Change Variables in the Discontinuous Growth Model*

Round (R)	TIME	Dissolution Transition (DT)	Dissolution Slope (DS)	Restoration Transition (RT)	Restoration Slope (RS)
Measurement occasion in the trust game	Linear change of MUs transferred in the formation phase (R1 to R5)	Difference in level of MUs transferred immediately following the trust violation (R6 vs R5)	Linear change of MUs transferred in the dissolution phase (R6 to R8)	Difference in level of MUs transferred immediately following the trust repair (R9 vs R8)	Linear change of MUs transferred in the restoration phase (R9 to R15)
1	0	0	0	0	0
2	1	0	0	0	0
3	2	0	0	0	0
4	3	0	0	0	0
5	4	0	0	0	0
6	4	1	0	0	0
7	4	1	1	0	0
8	4	1	2	0	0
9	4	1	2	1	0
10	4	1	2	1	1
11	4	1	2	1	2
12	4	1	2	1	3
13	4	1	2	1	4
14	4	1	2	1	5
15	4	1	2	1	6

Initially Level 1 change was examined by including only Level 1 predictors into the DGM. Next, the influence of between-individual factors was examined by including BPD and one of the other Level 2 predictors (preoccupied attachment, fearful attachment, self-protection beliefs, feelings of rejection) to account for differences in Level 1 change, while also controlling for the main effects of gender and cognitive ability. All tests conducted were two-tailed, and a criterion level of $p < .10$ was used for all cross-level interactions effects to account for insufficient power to detect cross-level interactions as a result of reduced parameter reliability in multilevel analysis (Snijders & Bosker, 1999).

Results

Descriptive data and intercorrelations

Table 8 presents the means, standard deviations, and intercorrelations for BPD, attachment styles, self-protective beliefs, and pre-game feelings of rejection. BPD had a medium positive association with fearful attachment style, self-protective beliefs, and pre-game feelings of rejection, and a small positive association with preoccupied attachment style. Trust formation has a small negative association with BPD. None of the other trust phases is significantly associated with any of the predictors.

To understand the mechanisms underpinning the trust behaviour patterns associated with BPD, a series of 4 models were created which, in addition to including BPD as a predictor, also examined the effects of individuals differences in preoccupied and fearful attachment styles, self-protective beliefs, and feelings of rejection. To be consistent with previous research (Abramov et al., 2020), the main effects of gender and cognitive reflective ability were also controlled for.

Table 8*Intercorrelations, Means, and Standard Deviations of Study Variables*

	1	2	3	4	5	6	7	8	
1 BPD	(.84)								
2 Preoccupied Attachment	0.25	-							
3 Fearful Attachment	0.34	0.35	-						
4 Self-Protective Beliefs	0.38	0.26	0.31	(.74)					
5 Feelings of Rejection	0.34	0.26	0.24	0.38	(.74)				
6 Trust _{Formation}	-0.18	0.03	-0.08	-0.11	-0.12	-			
7 Trust _{Dissolution}	-0.01	0.06	-0.07	0.06	0.00	0.34	-		
8 Trust _{Restoration}	-0.08	0.11	-0.04	0.07	-0.02	0.44	0.55	-	
	<i>Mean</i>	1.63	4.20	4.56	1.03	2.25	46.16	24.54	38.40
	<i>SD</i>	2.34	2.70	2.93	0.88	0.80	18.95	19.14	23.29
	<i>N</i>	234	233	233	234	234	234	234	234

Trust_{Formation} reflects average trust during the formation phase. Trust_{Dissolution} reflects average trust during the dissolution phase. Trust_{Restoration} reflects average trust during the restoration phase.

Spearman correlations are reported in the lower half. Alphas are reported on the diagonal. Values equal to or above |.22| are significant at $p < .01$ level. Values equal to or above |.18| are significant at $p < .05$ level. Values equal to or above |.16| are significant at $p < .10$ level.

Parameter estimates for the change in trust are provided in Table 9. Model 1 includes only BPD as a predictor of trust, while the remaining models reflect trust as a function of BPD and the following variables: preoccupied attachment style (Model 2), fearful attachment style (Model 3), self-protective beliefs (Model 4), and baseline feelings of rejection (Model 5).

Table 9

Discontinuous Mixed-Effects Growth Models Predicting Trust as a Function of BPD Trait Count (BPD), Attachment Style, Self-Protective Beliefs, and Feelings of Rejection, after Controlling for Gender and Cognitive Reflective Ability

	<u>Model 1</u>		<u>Model 2</u>		<u>Model 3</u>		<u>Model 4</u>		<u>Model 5</u>	
Moderator:	<i>None</i>		Preoccupied Attachment		Fearful Attachment		Self-Protective Beliefs		Feelings of Rejection	
	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>
Intercept	49.14***	2.02	48.73***	2.08	49.11***	2.11	48.35***	2.12	49.04***	2.11
Level 2 – Between Individual										
Gender (Female)	-4.30*	2.15	-4.15†	2.18	-4.25*	2.16	-4.34*	2.16	-4.26†	2.17
CRA	2.34*	1.04	2.36*	1.05	2.40*	1.03	2.33*	1.04	2.39*	1.04
BPD	0.17	1.49	0.09	1.71	0.56	1.92	-1.67	1.97	-0.43	1.75
Mod			-1.26	1.59	-1.01	1.62	0.81	1.71	1.31	1.65
BPD * Mod			0.98	1.46	-0.01	1.54	1.74	1.32	0.19	1.40
Level 1 – Within Individual										
Time	-0.08	0.57	0.11	0.60	-0.12	0.62	0.31	0.62	0.09	0.60
DT	-20.74***	2.40	-21.41***	2.52	-20.61***	2.59	-20.85***	2.58	-21.42***	2.52
DS	-0.76	1.16	-0.64	1.21	-0.73	1.24	-1.12	1.25	-0.64	1.23
RT	10.55***	1.86	9.49***	1.93	9.57***	2.00	10.21***	2.02	10.11***	1.97
RS	1.29**	0.40	1.32**	0.42	1.34**	0.43	1.16**	0.43	1.32**	0.42
Time * BPD	-1.12†	0.57	-1.20†	0.65	-1.14	0.74	0.08	0.75	-0.32	0.66
Time * Mod			1.29*	0.61	-0.04	0.62	-1.08†	0.65	-1.49*	0.62
Time * BPD * Mod			-0.63	0.56	0.07	0.59	-0.81	0.50	-0.44	0.53
DT * BPD	2.13	2.40	1.56	2.72	1.44	3.09	-1.03	3.13	-1.17	2.75
DT * Mod	-	-	-1.93	2.54	2.16	2.62	6.30*	2.72	6.18*	2.61
DT * BPD * Mod	-	-	2.37	2.34	-0.24	2.48	0.23	2.10	1.80	2.22
DS * BPD	2.33*	1.16	3.14*	1.31	3.56*	1.48	2.08	1.52	3.06*	1.34

DS * Mod	-	-	-1.29	1.22	-2.53*	1.25	-0.82	1.32	-1.47	1.27
DS * BPD * Mod	-	-	-0.76	1.13	-0.33	1.19	0.76	1.02	-0.32	1.08
RT * BPD	-3.83*	1.85	-6.38**	2.09	-6.46**	2.38	-4.25 [†]	2.45	-4.98*	2.15
RT * Mod	-	-	3.31 [†]	1.95	2.89	2.02	-0.38	2.13	1.40	2.05
RT * BPD * Mod	-	-	3.40 [†]	1.79	2.45	1.91	0.72	1.64	1.15	1.74
RS * BPD	0.43	0.40	0.48	0.45	0.44	0.51	0.05	0.52	0.55	0.46
RS * Mod	-	-	-0.10	0.42	0.19	0.43	0.34	0.45	-0.18	0.44
RS * BPD * Mod	-	-	-0.07	0.39	-0.13	0.41	0.26	0.35	-0.09	0.37
Variance Components	<i>Variance</i>	<i>SD</i>	<i>Variance</i>	<i>SD</i>	<i>Variance</i>	<i>SD</i>	<i>Variance</i>	<i>SD</i>	<i>Variance</i>	<i>SD</i>
Intercept	135.12	11.62	144.73	12.03	138.71	11.78	138.80	11.78	141.70	11.90
Time	15.68	3.96	15.41	3.92	15.90	3.99	14.63	3.82	14.07	3.75
DT	554.81	23.55	563.30	23.73	566.50	23.80	533.97	23.11	524.13	22.89
DS	39.01	6.25	38.53	6.21	35.50	5.96	36.75	6.06	36.80	6.07
RT	74.06	8.61	64.52	8.03	78.58	8.86	78.53	8.86	76.64	8.75
RS	13.75	3.71	13.96	3.74	13.91	3.73	13.81	3.72	14.01	3.74
Residual Error	561.01	23.69	561.32	23.69	563.21	23.73	561.20	23.69	560.76	23.68
Pseudo R ²	0.19		0.19		0.19		0.19		0.19	
df (<i>no. of Individuals</i>)	233		232		232		233		233	

BPD = Number of BPD traits reported on MSI-BPD. MOD = Moderator. CRA = Cognitive Reflective Ability. DT = Dissolution transition. DS = Dissolution slope. RT = Restoration transition. RS = Restoration slope. BPD, cognitive reflective ability, and all four moderators were *z*-standardized and centered at the sample mean.

[†]*p* < .10, **p* < .05, ***p* < .01, ****p* < .001, tests are two-tailed, *n*=234 participants, 3510 observations.

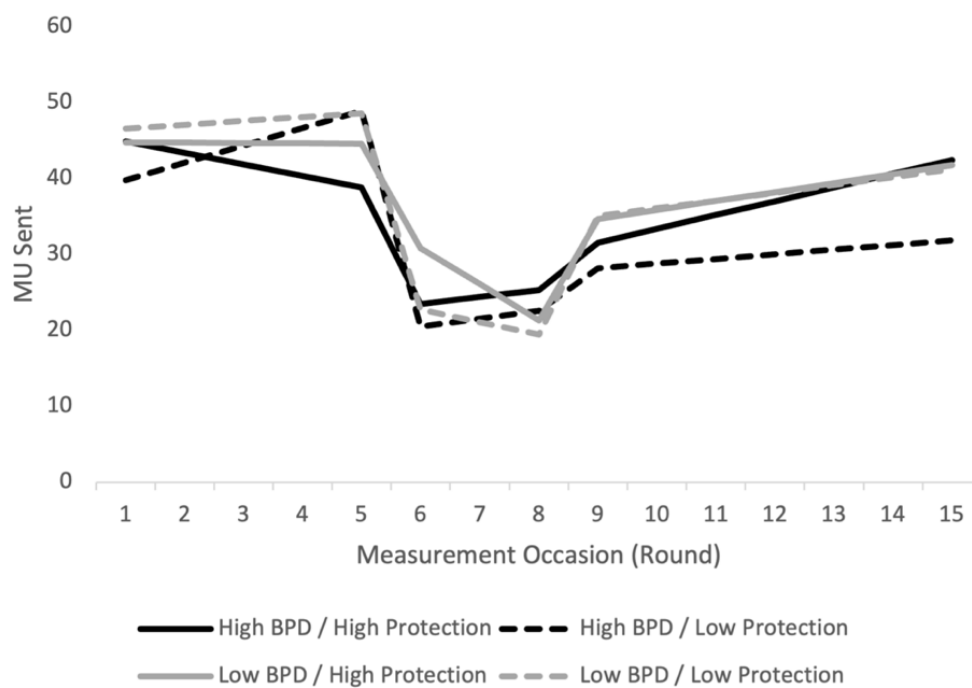
Formation slope

Growth in trust during the formation stage is represented by the estimates associated with Time. As seen in Model 1 in Table 9, individuals higher in BPD are associated with a decrease in trust growth during the formation stage (Model 1: $\text{est}_{\text{Time} * \text{BPD}} = -1.12, p < .10$). This effect is visualized in Figure 2 (Chapter 2). Individuals with higher BPD experienced a decreasing linear trend during the formation stage. This effect was present when controlling for preoccupied attachment style (Model 2: $\text{est}_{\text{Time} * \text{BPD}} = -1.20, p < .10$) and presented a trend towards significance after controlling for fearful attachment style (Model 3: $\text{est}_{\text{Time} * \text{BPD}} = -1.14, p = .12$). However, BPD no longer significantly predicted the rate of trust growth during the formation stage after controlling for self-protective beliefs (Model 4: $\text{est}_{\text{Time} * \text{BPD}} = .08, n.s.$) or baseline feelings of rejection (Model 5: $\text{est}_{\text{Time} * \text{BPD}} = -.34, n.s.$). Preoccupied attachment style predicted a positive linear trend in MUs transferred during formation after controlling for BPD (Model 2: $\text{est}_{\text{Time} * \text{Preoccupied}} = 1.29, p < .05$). However, preoccupied attachment was not found to significantly moderate the effect of BPD on the rate of MUs transferred during trust formation (Model 2: $\text{est}_{\text{Time} * \text{BPD} * \text{Preoccupied}} = -.63, n.s.$). Conversely, there was no statistical influence of fearful attachment on the growth of trust (Model 3: $\text{est}_{\text{Time} * \text{Fearful}} = -.04, n.s.$) nor evidence to suggest that fearful attachment moderates the relationship between BPD and the rate of MUs transferred (Model 3: $\text{est}_{\text{Time} * \text{BPD} * \text{Fearful}} = .07, n.s.$). Self-protective beliefs were associated with a gradual decrease in the number of MUs transferred, after controlling for the influence of BPD (Model 4: $\text{est}_{\text{Time} * \text{Protection}} = -1.09, p < .10$), and there was a trend towards self-protective beliefs moderating the effect of BPD on the MU transfer rate during trust formation (Model 4: $\text{est}_{\text{Time} * \text{BPD} * \text{Protection}} = -.81, p = .11$). As illustrated in Figure 3, higher levels of self-protective beliefs may exacerbate the negative effect of BPD on the rate of MUs transferred during the formation stage. Finally, baseline

feelings of rejection was significantly associated with declining trust during the formation phase (Model 5: $est_{Time * Rejection} = -1.49, p < .05$), but there was no evidence to suggest that baseline feelings of rejection moderated the relationship between BPD and the rate of MUs transferred (Model 5: $est_{Time * BPD * Rejection} = -.44, n.s.$).

Figure 3

Influence of BPD on Trust Growth Moderated by Self Protective Beliefs



Dissolution transition

BPD did not significantly predict the number of MUs transferred immediately following the trust violation, and this remained the case after accounting for attachment style, self-protective beliefs, and increased feelings of rejection in respective models (see Models 1-5 in Table 9). Self-protective beliefs (Model 4: $\text{est}_{\text{DT}} * \text{Protection} = 6.30, p < .05$) and baseline feelings of rejection (Model 5: $\text{est}_{\text{DT}} * \text{Rejection} = 6.18, p < .05$) were both found to predict a less pronounced decrease in the number of MUs transferred in response to the initial trust violation, after accounting for BPD. However, no significant three-way interactions between the dissolution transition, BPD, and each of the moderator variables were observed (see Models 2-5 in Table 9).

Dissolution Slope

As seen in Table 9, BPD was found to significantly predict a faster rate of growth in the number of MUs transferred during the dissolution phase (Model 1: $\text{est}_{\text{DS}} * \text{BPD} = 2.33, p < .05$). See Figure 2 for a visual representation (Chapter 2). Significance was maintained when attachment style and baseline feelings of rejection were accounted for in respective models (see Models 2, 3, and 5 in Table 9). When self-protective beliefs were accounted for, BPD no longer significantly influenced trust growth during the dissolution phase (see Model 4 in Table 9). Only fearful attachment style significantly predicted the rate of MUs transferred during the dissolution phase after accounting for number of BPD symptoms and had a negative effect on growth (Model 3: $\text{est}_{\text{DS}} * \text{Fearful} = -2.53, p < .05$). No significant three-way interactions between the dissolution slope, BPD, and each of the moderator variables were observed (see Models 2-5 in Table 9).

Restoration transition

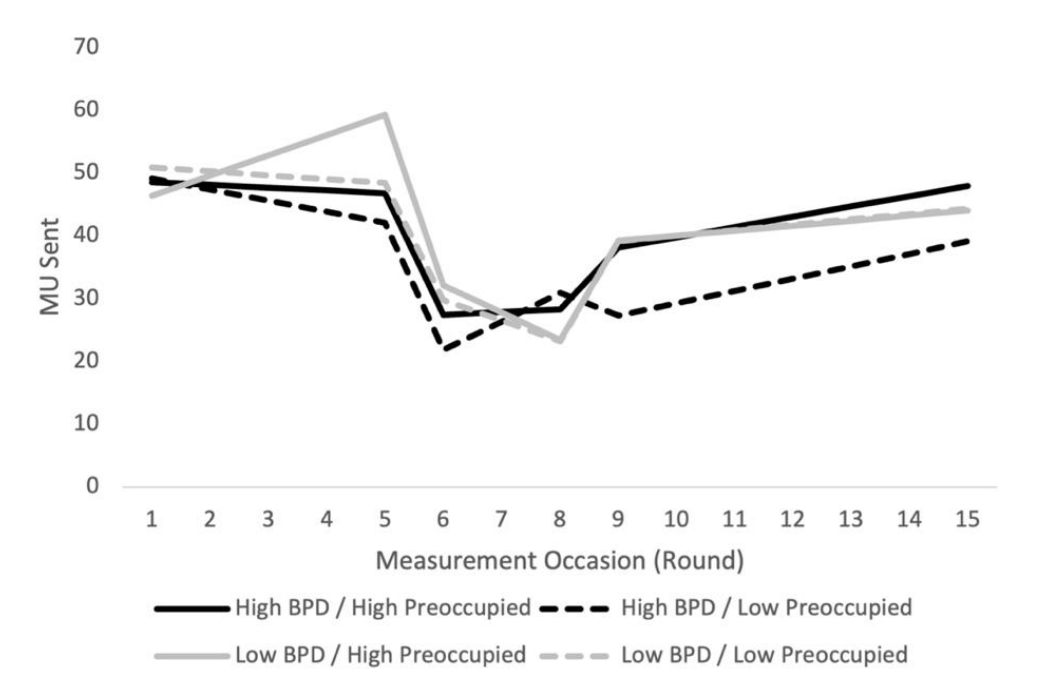
BPD was found to significantly predict how many MUs were transferred immediately following the first instance of trust repair (Model 1: $est_{RT} * BPD = -3.83, p < .05$), by reducing the size of the investment. After controlling for each of the moderator variables, this effect continued to be significant. Only preoccupied attachment style was found to predict the size of the restoration transition after accounting for BPD, but in contrast to BPD, the effect was positive (Model 2: $est_{RT} * Preoccupied = 3.31, p < .10$). Preoccupied attachment was also found to significantly moderate the effect of BPD on the rate of MUs transferred at the restoration transition (Model 2: $est_{Time} * BPD * Moderator = 3.40, p < .10$). As illustrated in Figure 4, results suggest that higher levels of preoccupied attachment counteract the negative effect of BPD on the rate of MUs transferred at the restoration transition.

Restoration slope

BPD did not significantly predict the rate of growth in the number of MUs transferred during the restoration phase, and this remained the case after accounting for attachment style, self-protective beliefs, and baseline feelings of rejection in separate models (see Models 1-5 in Table 9). Additionally, none of the moderator variables was found to predict trust growth during restoration after controlling for BPD, and no significant three-way interactions between the restoration slope, BPD, and each of the moderator variables were observed (see Models 2-5 in Table 9).

Figure 4

Influence of BPD on Trust Growth Moderated by Preoccupied Attachment Style



Discussion

Summary

The current study used an experimental TG to examine whether attachment style, self-protective beliefs, and feelings of rejection explain the paradoxical relationship between BPD and trust (Abramov et al., 2020). Results suggest that the relationship between borderline pathology and the trajectory of trust growth when it is forming appears to be influenced by a number of these variables. Endorsing the belief that one needs to take action to protect oneself appears to underlie and possibly reinforce, the effect of BPD on the way trust is

formed and potentially dissolved. First, when such beliefs were controlled for, the decline in trust during the formation phase associated with BPD was no longer significant. Second, there was evidence of a non-significant trend that the gradual decline in trust during the initial phase of the interaction found to be associated with borderline pathology, appeared to become more pronounced as self-protective beliefs increased. Third, self-protective beliefs were associated with a smaller decline in funds sent after the initial trust violation. Moreover, the paradoxical growth in trust in response to multiple, consecutive trust violations that was associated with borderline pathology no longer reached significance when level of self-protective beliefs was taken into account.

Given the relationship between BPD and the expectation of betrayal and abuse by others (Barazandeh et al., 2016; Bhar et al., 2008), it is possible that individuals with a high number of BPD traits entered the game already believing that the other player was untrustworthy and likely to betray. These beliefs likely activate the use of self-protecting behaviours, reflected in the increasingly smaller amounts invested during the formation phase. This is also in line with previous research where, compared to HCs, individuals with BPD tended to be more pessimistic when predicting TG outcomes even in the absence of any feedback regarding trustee reciprocity (Unoka et al., 2009). The study authors suggested the lowered expectancies were related to more general beliefs about the trustworthiness of others, rather than beliefs specific to their TG partner. Indeed, there is evidence that BPD trait count is not associated with perceptions of TG partner trustworthiness or fairness (Abramov et al., 2020), and individuals with BPD have not been found to differ from HCs in the accuracy of their appraisals of the fairness of trustee reciprocity (Franzen et al., 2011).

This gives weight to the notion that the mistrustful behaviours exhibited during the beginning of the TG despite engaging with a cooperative partner, are self-protective responses shaped by past experiences of betrayal, rather than reactions to the actual trustworthiness (reciprocity) of the partner or systematic differences in how reciprocity levels are appraised as indicators of trustworthiness. As such, holding self-protective beliefs may override the experience of an objectively cooperative partner, and it is even possible that the partner's repetitive cooperative exchanges make the anticipated betrayal more salient. Waiting for the 'inevitable' breach of trust with a new, cooperative partner, much like waiting for the other shoe to drop, may then be associated with taking self-protective action to first mitigate the risk by investing defensively, before then trying to salvage the relationship with a less negatively reactive response to the initial trust violation. That is, perhaps the mistrustful stance observed during the formation phase was a conscious strategy to obtain evidence that the other will betray, and the initial betrayal having validated the expected 'rules of engagement', allowed the high BPD trait individual to attempt to reengage the trustee by making progressively larger investments. In line with this, De Panfilis et al (2019) found that in other economic games, BPD patients were more likely than controls to punish their partners when receiving fair offers, but this was not the case when they received unfair offers.

As observed with self-protective beliefs, existing feelings of rejection on trust growth during the initial phase of the game appears to exert a suppressant effect on the rate of trust formation, which potentially explains the declining trust associated with high levels of BPD traits. While rejection sensitivity was not explicitly measured, these results were consistent with previous findings that sensitivity to rejection may mediate the relationship between BPD and negatively biased appraisals of trustworthiness (Miano et al., 2013; Richetin et al., 2018). Greater feelings of rejection were also associated with a less pronounced fall in trust in

response to the initial trust violation which, like self-protective beliefs, may reflect an expectation of and therefore less reactivity to rejection or betrayal by the other.

While self-protective beliefs and feelings of rejection appear to reinforce some of the trust patterns observed in individuals with high levels of BPD traits, the findings on attachment insecurity suggest that the influence of borderline pathology on trust patterns occurs *despite* the contrasting effects of attachment insecurity. For example, while borderline pathology was associated with greater mistrust when trust was forming and in response to the initial trust repair effort, a preoccupied attachment style was associated with a faster rate of trust formation, and a more generous response to initial repair. This is in line with previous empirical findings that attachment anxiety is associated with increased interpersonal anxiety in response to affiliative overtures from a potential close other, a strong preference to making affiliative overtures, and a preoccupation with reciprocity (Bartz & Lydon, 2006, 2008; Fett et al., 2016). In fact, as preoccupied attachment increased, the negative effect of BPD on trust behaviours in response to the initial trust repair became less pronounced, suggesting preoccupied attachment style may reflect a more socially normative style of relating, whereby cooperation is rewarded and uncooperativeness is punished, as observed in general populations (Fulmer & Gelfand, 2015; Korsgaard et al., 2018).

In contrast, whereas borderline pathology was associated with a paradoxical growth of trust in response to multiple, consecutive trust violations, individuals reporting higher levels of fearful attachment responded by gradually reducing trust. This may have reflected the importance of reciprocity among those reporting greater levels of attachment anxiety, although curiously, it was fearful rather than preoccupied attachment that reached significance. However overall, these findings suggest that attachment insecurity does not appear to explain the paradoxical increase in trust behaviours in high BPD trait individuals.

These findings suggest that despite the theoretical and empirical links between BPD and preoccupied and fearful attachment styles, attachment insecurity was not found to drive the unusual trust patterns observed in the high BPD trait individuals and may in fact have tempered these patterns.

Limitations & future directions

The current study has several limitations. First, a clinical sample of individuals with BPD was not used. Although relational disturbances in nonclinical samples of people with borderline personality traits are almost as acute as in clinical samples (Tolpin et al., 2004), it cannot be assumed that individuals with a clinical diagnosis would respond similarly. It is notable however, that despite the modest proportion of participants endorsing a clinically relevant number of BPD traits, there was nevertheless sufficient evidence to suggest these variables play a role in the relationship between BPD and trust growth.

Second, the use of anonymous one-time encounters may not have been sufficient to evoke attachment-salient responses. Previous researchers examining trust have attempted to trigger the underlying attachment system by using romantic partners and engaging them in relationship threatening dialogue (Miano, Fertuck, et al., 2017) or through the administration of oxytocin (Bartz, Simeon, et al., 2011). Although a trust violation was created with the view that it might elicit an attachment relevant behavioural response, it cannot be assumed that it was sufficient. However, by using unknown partners, the current study offers an important insight into how borderline pathology might influence how new relationships are developed. Finally, a simulated trustee was used rather than an actual human being. While this methodology allowed for the systematization of trustee reciprocity, there is prior evidence that the effect of attachment on cooperation in social contexts was only observed

when the participants were partnered with a human rather than a computer (Taheri et al., 2018). While efforts were made to create the illusion that participants were playing against a human in the study, the possibility remains that the deception was not effective. It is possible that effects may be artificially suppressed.

This research augmented previous TG studies by examining how attachment style, self-protective beliefs, and feelings of rejection accounted for or modified the relationship between BPD and trust behaviours using an economic exchange game. The cross-disciplinary methodology provides a prototype for future research looking at interpersonal disturbances as dynamic, dyadic processes. The sample provided a conservative estimate of the complex interaction between BPD traits and multiple sources of interpersonal disturbances on trust formation, dissolution, and restoration in a discrete social exchange. It is important for future research to replicate these findings in a clinical sample, as well as examine trust processes for individual with BPD within other dyads such as friends, family, and romantic unions, to determine whether the findings apply in these more specific and ‘higher stakes’ relationships.

In this study trust behaviours were quantified to explore how various features of the borderline personality might influence the relationship between BPD and trust, to make inferences about what social cognition processes might be altered in individuals with a borderline presentation. The most compelling finding was that the presence of self-protective beliefs and feelings of rejection may have a detrimental effect on how individuals with BPD traits for trust in interpersonal exchanges. A potential next step would be to augment these findings by using qualitative methodology to examine how individuals make sense of their own and their partner’s behaviour in such an interaction (Sharp, Ha, et al., 2011). This process, known as mentalization (Allen et al., 2008; Bateman & Fonagy, 2004), has been

found to be compromised among people with BPD (for reviews see Dinsdale & Crespi, 2013; Jeung & Herpertz, 2014; Lazarus et al., 2014; Mitchell et al., 2014; Richman & Unoka, 2015). Meta-analytical findings suggest the deficits observed in BPD are not decoding impairments, but rather relate to the process of reasoning (Németh et al., 2018), that is, reasoning about others' mental states in order to explain or predict behaviour (Sabbagh, 2004). Examining reasoning would allow researchers to better understand how individuals with BPD experience the mutability of interpersonal exchanges and explore in greater depth how feelings of rejection and beliefs about the need to protect against betrayal by others may compromise the development of trust.

Conclusions

This study built on previous research using DGM with a TG to examine the how attachment style, self-protective beliefs, and feelings of rejection might underpin or modify the anomalous trust patterns associated with BPD traits. The findings reveal that the slower rate that trust is formed among high BPD trait individuals may be underpinned or accompanied by feelings of rejection and beliefs that others will betray so one should act preemptively to protect oneself. Attachment insecurity appears to have an influence on trust that is in direct contrast to the effect that BPD exerts. Finally, BPD remained a robust predictor of faster trust growth in response to the trust violation, only partially being explained by self-protective beliefs, suggesting that there is something unique to the borderline experience that creates paradoxical ways of responding in trust-based situations. Regrettably, this manner of interacting is likely to elicit and compound the interpersonal difficulties such individuals face. The findings highlight the complex nature of borderline relational disturbance, and the need for research that can assess and quantify the internal experience of individuals with BPD to explain the aberrant style of social exchange associated with this personality disorder.

Chapter 4: Social-cognitive reasoning for decisions made in a trust game for individuals with high or low borderline personality disorder trait count

This chapter has been submitted for publication. The chapter is identical to the submitted manuscript except for table numbers (Tables 1, 2, & 3 renamed 10, 11, & 12, respectively), which have been altered to ensure uniformity in formatting across the thesis.

Abramov, G., Miellet, S., & Deane, F. P. (2022). Social-cognitive reasoning for decisions made in a trust game for individuals with high or low borderline personality disorder trait count. *Manuscript submitted for publication.*

Abstract

Borderline personality disorder (BPD) is associated with paradoxical trust behaviours in interpersonal exchanges, but little is known about the thought processes behind these behaviours. The current study aims to understand the social-cognitive reasoning for decisions made in a trust game before, during, and after a trust violation based on number of BPD traits endorsed. Thirty-two university students reporting either a high (7-10) or low (0-2) number of BPD traits played a 15-round trust game in which cooperation by the other player was varied to create three phases of trust: formation, dissolution, and restoration. Afterwards participants provided a free form response to explain the decisions they and the other player made in selected rounds. During trust formation, the high BPD trait group more frequently attributed negative characteristics to the other player and described their own behaviour as a tit-for-tat strategy despite reasoning that their partner was responding with positive reciprocity. Overall, the high BPD trait group was also more likely to say that they did not know why they or the other player made the decisions they did, potentially reflecting deficits in social-cognitive reasoning. Interventions which focus on improving the capacity and accuracy of social-cognitive reasoning may address the trust-related interpersonal difficulties associated with BPD. A possible limitation is that social-cognitive reasoning was elicited after completing the trust game so may not reflect in-the-moment reasoning. Future research could examine how prompting reasoning prior and/or during the exchange influences trust behaviours.

Introduction

Maladaptive interpersonal functioning is a key diagnostic indicator of borderline personality disorder (BPD) (APA, 2013), with those afflicted experiencing intense, unstable, ruptured, and highly conflicted relationships (Bouchard et al., 2009; Clifton et al., 2007; Lazarus et al., 2020). Impairments in social-cognitive functioning are proposed to underpin the interpersonal difficulties in BPD and are well-documented among this population (for reviews see Dinsdale & Crespi, 2013; Lazarus et al., 2014; Németh et al., 2018; Roepke et al., 2012; Sharp, 2014). Navigating social exchanges successfully is largely dependent on the social-cognitive capacity to accurately identify the intentions, thoughts and emotions of others and oneself, variously referred to as mentalizing and reflective functioning (Bateman & Fonagy, 2004), and Theory of Mind (ToM) (Premack & Woodruff, 1978). Meta-analytical research suggest it is impairments in reasoning processes rather decoding that underlie the deficits associated with BPD (Németh et al., 2018). Individuals with BPD have been found to have impairments in self-reported mentalization as well as task-based perspective-taking to infer the mental and/or emotional state of others (Bora, 2021). Hypermentalizing is a style of mentalizing in which an individual makes inferences about another person's mental states that far exceed what could be justified by the observable data (Sharp et al., 2013). Hypermentalization is associated with BPD in adults (Andreou et al., 2015; Vaskinn et al., 2015) and adolescents (Cortés-García et al., 2021; Penner et al., 2020; Quek et al., 2019; Sharp et al., 2013; Sharp, Pane, et al., 2011; Somma et al., 2019).

One area of interpersonal functioning identified to be aberrant in BPD and potentially linked to social cognition deficits is that of trust behaviour (Masland et al., 2020). Economic exchange games provide a framework to examine social cognition processes in 'real life' interactions, providing the benefits of greater ecological validity within a controlled

environment. The trust game (TG: Berg et al., 1995) has been used to examine decision making in trust-based social exchanges. In a typical TG, a ‘trustor’, also referred to by the label ‘investor’, decides what proportion of a monetary allocation to send to the other player, the ‘trustee’, with the objective being to receive a larger sum in return, but the risk being that the trustee will keep most or all of it. Trust is operationalized as the proportion of the monetary allocation sent by the trustor. TG research has found that BPD is associated with reduced cooperation by trustees (King-Casas et al., 2008) and reduced trust by trustors (Abramov et al., 2020; Liebke et al., 2018; Niedtfeld & Kroneisen, 2020; Unoka et al., 2009).

In our previous work we used a TG programmed to create three trust phases: trust formation (trustee is cooperative and investment returns are equal to or greater than the amount invested), trust dissolution (trustee violates trust by keeping most or all of the amount invested), and trust restoration (trustee returns to cooperative play as during formation). We sought to understand the relationship between trust behaviours and BPD trait count during each of these phases. To our surprise we found that individuals with a high number of BPD traits appear to exhibit paradoxical interpersonal trust behaviours in that they responded with increasing mistrust during the initial interactions with a cooperative partner, and increasing trust in response to a drop in partner-cooperation (Abramov et al., 2020). In a follow-up study we found that endorsing the self-protective belief that others will betray and that one should pre-emptively act to protect oneself, and feelings of rejection prior to commencing the TG, each explained the association between borderline pathology and mistrust during the initial interaction when the partner was cooperative (Abramov et al., 2022). Moreover, when self-protective beliefs were considered, BPD trait count no longer predicted the paradoxically more trusting behaviour in response to the trust violations. These findings highlight the need

to augment the existing research with an examination of the social-cognitive reasoning related to trust behaviours in individuals with borderline features.

Social-cognitive processes are believed to be engaged when individuals make strategic interpersonal decisions (Frith & Singer, 2008; Hampton et al., 2008; Lee & Harris, 2013), such as would occur in trust salient contexts. In the broader literature, trust behaviours have been linked to social-cognitive processes such as perspective taking (Fett et al., 2014; Sun et al., 2021), theory of mind (Derks et al., 2015), and reasoning about the intentions of others (Sharp, Ha, et al., 2011). To date, none of the studies examining anomalies in trust and cooperation associated with borderline pathology have explicitly examined social-cognitive reasoning although neural correlates have been identified to explain uncooperative behaviour in trustees with BPD. In the first study of its kind in the BPD literature, King-Casas and colleagues (2008) used a 10-round TG in which trustees with or without BPD played against trustors (non-BPD controls). They found that as trustee cooperation faltered, trustors signalled mistrust by sending less money, but there were remarkable differences in how trustees responded. Controls responded by increasing the amount they repaid, in effect ‘coaxing’ the trustors’ ongoing cooperation, while the BPD group responded by further reducing the amount reciprocated, leading to a breakdown in cooperation. Neuroimaging revealed that unlike the controls, the BPD group did not exhibit a neural response in the anterior insula, an area associated with norm violations. The authors suggested that individuals with BPD may have a diminished capacity to interpret the expectations or motives behind their partners behaviour, and therefore miss the opportunity to take corrective action to preserve the relationship (King-Casas et al., 2008).

In contrast, Franzen and colleagues investigated theory of mind performance on trust behaviours in BPD by examining the effect of both emotional cues and fairness, linked to repayment rates, on trust behaviours (Franzen et al., 2011). Compared to controls, BPD patients titrated their level of trust (size of investment sent) to their partners *actual* level of fairness in the presence of fairness-incongruent facial expressions. The authors suggested that the superior abilities of the BPD group to infer the intentions of their partner may have reflected the use of explicit-controlled processing rather than automatic processing that controls may have used (Franzen et al., 2011). Post-hoc analysis of response times revealed slower response times in the BPD group which may reflect increased deliberation resulting in more rational responses. Importantly, Franzen et al.'s (2011) participants played single round games against multiple partners whereas in King-Casas et al. (2008), it was an iterated TG meaning players are likely to have been taking into consideration their partner's behaviour across multiple rounds, potentially placing more strenuous demands on mentalization capacity.

While diminished or impoverished mentalization may be a possible mechanism underpinning trust behaviour in BPD, viewed through a hypermentalization lens, the tendency of King-Casas et al.'s BPD group to act in a way to deepen the rupture may reflect a misattribution of hostile intent to the trustor's social signalling. Supporting this hypothesis is the finding that in comparison to controls, trustors with BPD both send less money and have more negative predictions of TG outcome even in the absence of any feedback regarding trustee reciprocity (Unoka et al., 2009). These findings may reflect a disposition towards perceiving others as untrustworthy, which is consistent with findings of a negative trust bias associated with BPD (for review see Fertuck et al., 2018; Masland et al., 2020), as well as beliefs that others will betray, exploit and deceive (Barazandeh et al., 2016; Bhar et al.,

2008). Immutable expectations of negative treatment may also explain our own research where we found that BPD trait count was associated with growing mistrust when interacting with a cooperative trustee (Abramov et al., 2020), but this effect was no longer significant when self-protective beliefs or pre-game feelings of rejection were taken into account (Abramov et al., 2022).

BPD has been linked to both rejection-related (Lobbestael & McNally, 2016) and hostility related (Arntz et al., 2011; Barnow et al., 2009; Lobbestael & McNally, 2016) interpretation biases. People with BPD are more likely to feel excluded by others even when there is objective evidence that they are included (De Panfilis et al., 2015; Domsalla et al., 2014; Renneberg et al., 2012; Staebler et al., 2011; Weinbrecht et al., 2018). Rejection sensitivity, or the anxious expectation that one will be rejected, is associated with greater hostility in response to perceived rejection (Ayduk et al., 2008). In addition, borderline pathology is linked to increased feelings of anger and hostility in response to perceived or experienced rejection (Berenson et al., 2011; Hepp, Lane, et al., 2018; Lazarus et al., 2018; Richmond et al., 2020; Scott et al., 2017). It is possible that individuals high in BPD traits approach the TG already feeling rejected and perceive rejection regardless of trustee cooperativeness, and therefore express hostility by sending smaller amounts for investment.

Such findings support the argument that examining anomalous trust behaviours through a social-cognitive lens is an important next step in BPD trust research. A suitable protocol was developed by Sharp and colleagues (2011) who used a TG with a community sample of boys to examine whether social cognition deficits underpinned differential trust behaviours associated with externalizing behaviour problems. They asked participants to evaluate both their own and their partners intentions during a TG, and found that in boys with

externalizing problems, untrusting and untrustworthy behaviours were associated with both evaluating the other player's intentions as unfair and their own hostile intentions, rather than general theory of mind impairments (Sharp, Ha, et al., 2011). The current study uses a similar method to Sharp et al. (2011) by asking individuals with a high number of BPD traits to explain both their own and their partners actions during a TG in which repayments are modified to elicit a violation of trust by the partner, and later, efforts to repair trust by the partner (Abramov et al., 2020). This will allow examination of how individuals with borderline pathology interpret or provide reasoning about their partner's behaviours and rationalize their own behaviours during various phases of the game. Specifically, the initial phase where the partner reciprocates (trust formation), the phase in which the partner fails to reciprocate (trust violation and dissolution), and the phase in which there is a return to reciprocity by the partner (trust repair and restoration). The current study utilizes an economic game framework to examine the social-cognitive reasoning associated with trust behaviours for individuals with a high versus low BPD trait count.

Method

The study was approved by the University of Wollongong ethics committee (HE2017/253). All participants provided informed consent. This data is part of a larger research project and focuses on data not previously published (Abramov et al., 2020, 2022).

Participants

Participants ($N = 32$; 69% female; $M = 18.94$ years, $SD = 1.11$ years) were drawn from a sample of undergraduate students from a large Australian university ($N = 234$; 64% female; $M = 20.87$ years, $SD = 5.66$ years) who had been invited to take part in an online study looking at the relationship between economic decision making and personality

variables in exchange for course credit. To determine BPD trait count, participants completed the McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD; Zanarini et al., 2003), a 10-item self-report screening instrument with satisfactory internal consistency, validity, test-retest reliability, and in individuals aged 18 to 25, excellent sensitivity (.90) and specificity (.93) for the diagnosis of DSM-IV BPD (APA, 1994). The MSI-BPD demonstrated very good internal consistency in the full sample from which participants in the current study were drawn (Cronbach's $\alpha = .84$, $N = 234$). The MSI-BPD uses a conservative cut-off of 7/10 to reflect caseness for BPD. Sixteen participants met caseness in the current sample. A cut-off score of ≤ 2 was used to identify 177 participants with a low BPD trait count, as three or more symptoms has been linked to clinically relevant levels of BPD pathology (CLPS; Gunderson et al., 2011). For the current study we randomly matched the 16 participants with a high BPD trait count with 16 age- and gender-matched participants with a low BPD trait count to create two groups hereby referred to as high BPD and low BPD.

Experimental Procedure

The experiment was conducted entirely online and participants who had signed up to participate were provided with a link to access information about the study, consent forms, and study materials as described next. The study consisted of three parts: (a) self-report questionnaires, (b) trust game, (c) post trust game questions about own (self) and game partner's behaviour. Upon registering and giving consent, participants provided demographic data and completed a series of questionnaires as part of the larger study (see Abramov et al., 2020, 2022 for full list of measures). Instructions were provided on how to play the TG and participants were tested to ensure an adequate understanding of the game before being permitted to proceed. Participants were all assigned the role of trustor and advised they would

be partnered anonymously with another study participant to play the TG but were in fact paired with a computer programmed trustee, this being a common form of deception in BPD TG studies. This form of deception has been used in other studies using the same protocol in a BPD population (e.g., Liebke et al., 2018; Unoka et al., 2009). Participants were not advised how many rounds they would play as defection rates have been found to increase towards the end of a social exchange (Murnighan, 1981).

The current study used an iterated 15 round version of the TG (Berg et al., 1995). At the start of each round, trustors were allocated 100 monetary units (MUs) from which they could send any amount (0-100 MU) to the trustee for investment. After receiving the transferred amount multiplied by a factor of 3, the trustee could then return a proportion ranging from 0-100% to the trustor. Following each round, the participant was provided with a summary indicating how much they invested, how much was repaid, and the final sum each party ended up with for that round. These values were reset for each new round.

To examine how trust might form, dissolve, and restore, we modified trustee reciprocity to create a trust violation and a trust repair. During rounds 1-4 and 8-15 inclusive, trustee repayments were randomized to fall between 34-44% of the tripled investment, providing the trustor with a return the equivalent of the original investment plus up to 32% profit (range 0-32%). During rounds 5-7 inclusive, trustee repayments were randomized to fall between 0-10% of the tripled investment, providing the trustor with a return the equivalent of losing from 70% to 100% of their original investment⁴. The rationale for choosing these reciprocity ranges and a more detailed description of the TG protocol is presented in our original study (Abramov et al., 2020). Based on this repayment schedule,

⁴ Due to rounding to the nearest whole number, investments of 1 MU did not incur a loss.

investments can be divided into three distinct trust phases: *formation* (rounds 1-5), *dissolution* (rounds 6-8), and *restoration* (rounds 9-15). As participants see the summary at the end of each round, trust rates lag trustee reciprocity rates by 1 round (e.g., trust round 4 would be based on reciprocity round 3).

After completing the TG, participants were informed that they would be asked a series of questions about the choices they and the other player made. For selected rounds (1, 3, 5, 6, 7, 8, 9, 12, and 15) participants were presented with a prompt describing the action taken and asked to provide reasoning, both for their own investment (e.g., *In round 5 you sent \$60. Why did you choose to send this amount?*) and their partner's repayments (e.g., *In round 5 the other player repaid \$75. Why do you think they chose to repay this amount?*). Questions were structured following the natural order of the game (e.g., round 1 self, round 1 other, round 3 self, round 3 other...), and questions were presented one at a time with a prompt to type the answer before being permitted to click ahead to the next question. Table 10 outlines what trust phase the self or partner social-cognitive reasoning reflects by game round and trustee reciprocity proportions. In sum, each participant was asked to provide nine responses pertaining to their own behaviour and nine responses pertaining to their partner's behaviour.

Table 10

Game Round, Trustee Reciprocity and the Associated Phase of Trust for Self and Partner Social-Cognitive Reasoning

Round	Trustee Reciprocity	Reasoning – Self	Reasoning - Partner
1	34-44%	Formation	Formation
2	34-44%	-	-
3	34-44%	Formation	Formation
4	34-44%	-	-
5	0-10%	Formation	Dissolution
6	0-10%	Dissolution	Dissolution
7	0-10%	Dissolution	Dissolution
8	34-44%	Dissolution	Restoration
9	34-44%	Restoration	Restoration
10	34-44%	-	-
11	34-44%	-	-
12	34-44%	Restoration	Restoration
13	34-44%	-	-
14	34-44%	-	-
15	34-44%	Restoration	Restoration

Results

A total of 576 responses were coded. Participants expressed a variety of reasons, which will also be referred to as attributions, for their own and their partner's behaviours in the game. A content analysis of the responses identified 9 categories of social-cognitive reasoning to explain behaviour: (a) fairness, (b) prosocial, (c) positive reciprocity, (d) trust, (e) mistrust, (f) negative reciprocity, (g) personal gain, (h) negative partner traits, (i) negative affect. These are listed in Table 11 along with verbatim examples of each. Most of these

categories were also identified in other economic game studies in which participant and/or partner motives were examined (Chita-Tegmark, 2018; Dreber et al., 2014; Sharp, Ha, et al., 2011). In addition, on occasion participants stated they did not know why they or their partner had taken the action they did, which was categorized as (j) 'don't know'. Finally, all other responses including blank responses, playing randomly, nonsensical, or descriptive responses were categorized under (k) other.

Participant responses were coded based on the descriptions and examples provided in Table 11. All demographic and psychometric data was removed prior to coding. Responses were assigned a code of one for each type of reason provided. A single response could receive a code for multiple categories (e.g., "They were selfish and wanted the money for themselves" was coded as both personal gain and negative partner traits). A random selection of 58 self-responses and 58 partner-responses were blind coded by a second independent coder to check the reliability of the coding. Kappa coefficients were calculated to assess interrater reliability. The Kappa coefficient for self-statements was 0.96 and for partner-statements was 0.84 indicating very high interrater reliability.

Table 11*Categories of Social-Cognitive Reasoning for Self and Partner Trust Game Behaviours*

Category	Definition	Own Behaviour	Partner Behaviour
Fairness	Fairness, equality, mutual benefit, or unfairness explicitly mentioned.	<ul style="list-style-type: none"> - I thought it was fair. - It was a fair and even amount. - To benefit us both. 	<ul style="list-style-type: none"> - Coming out even. - Because I wasn't fair. - To be fair but still profit.
Prosocial	Cooperating not conditional on the other player's behaviour.	<ul style="list-style-type: none"> - I wanted to be nice. - Still wanted to remain giving. - Because it was the first found and I wanted to be generous. 	<ul style="list-style-type: none"> - Must've been feeling generous! - Bc there a good m8. - Because he decided to be nice.
Positive reciprocity	Cooperation in response to perceived cooperation by the other player.	<ul style="list-style-type: none"> - Because I was satisfied with how much they repaid the previous round and wanted to closely match it. - They were becoming less greedy therefore so was I. 	<ul style="list-style-type: none"> - Slightly above what I sent. - They saw I was giving generous amounts so started to increase their payments. - Because I was increasing payments.
Trust	Trust/trustworthiness or testing the other player or attempts to elicit greater cooperation explicitly mentioned.	<ul style="list-style-type: none"> - To see how giving they would be. - Trying to make the investment larger with a larger pay in. - I wanted to gain their trust so they would start giving me more money. 	<ul style="list-style-type: none"> - Showing they are willing to give slightly more. - They had realised that being greedy wasn't the way to go. - To give me back more and hope I would return with more. - Make the investment money higher. - They trusted me again but were selfish and wanted the money.
Mistrust	Lack of trust/trustworthiness or focus on safety.	<ul style="list-style-type: none"> - Started off small. - I thought it was a safe amount. - Declining trust in the other person. - Wanted to save on money. 	<ul style="list-style-type: none"> - To ensure he didn't lose too much but also to pay me back. - They didn't trust me. - Smaller amount. - Was more scared to spend money.
Negative reciprocity	Reducing or ceasing cooperation in response to perceived lack of cooperation by the other player, tit-for-tat strategy.	<ul style="list-style-type: none"> - Because they weren't giving me shit. - I was copying their low payment amount. - Because they only sent me \$5. 	<ul style="list-style-type: none"> - They weren't impressed with the amount I got. - Because I sent a low amount. - Because I did it first. - Because this game became more selfish_ and neither of us wanted to give money.

		<ul style="list-style-type: none"> - Because the player sent 0 before so I wanted to play dirty. - Because they were being greedy. 	
Personal gain	Increasing personal gain at the expense of the other.	<ul style="list-style-type: none"> - I wanted to win. - Need to regain capital. - Because I didn't want him to gain more money. - Because I didn't want to invest with them anymore. 	<ul style="list-style-type: none"> - They were selfish and wanted the money for themselves. - They were trying to get as much money as possible. - They don't care about sending money to strangers.
Negative partner traits	Negative characteristics attributed to partner.	<ul style="list-style-type: none"> - Because they were being rude) and giving me no money. - Because the other player was being stingy. - Because they were being greedy. 	<ul style="list-style-type: none"> - Cheap dog. - Rude. - Because the player is being mean. - I thought it was quite greedy to not give anything at all.
Negative affect	Reference to own negative affect, or attribution of negative affect to partner.	<ul style="list-style-type: none"> - I wasn't happy with the amount I was receiving back so they didn't deserve money. - I felt bad. - Because they annoyed me so I only sent back something small. - Angry. 	<ul style="list-style-type: none"> - Mad because I paid low. - They were annoyed I hadn't sent a large amount. - They were unhappy.
'Don't know'	Do not know why they or their partner behaved that way.	<ul style="list-style-type: none"> - I don't know. - Not sure. 	<ul style="list-style-type: none"> - Not sure. - No clue? - ?
Other	Miscellaneous responses including statements describing the behaviour or the game without reasoning, statements that the behavioural choices were random, blank responses, and nonsensical responses.	<ul style="list-style-type: none"> - Because it was the first round. - Hadn't sent that amount. - Random number. - So they could buy a pair of shoes. - Because. - Was increasing by intervals of 10 each round. 	<ul style="list-style-type: none"> - Yeah. - I just did. - Oh my god. - Because I didn't send any (when the trustor sent nothing, then the trustee would have not money to return)

Table 12 presents the frequency of responses for each type of reasoning category across each phase of the game to explain the participants' own and their partners behaviour, as per BPD trait level (high versus low). In response to 576 prompts (16 per participant) a total of 632 codes were assigned due to some responses meeting the coding requirements for multiple social-cognitive reasoning categories. Importantly, to ensure that frequencies reflected the number of participants that mentioned a category rather than the number of times a category was mentioned, a maximum frequency value of one was given for each participant for self or partner reasoning during each phase of the TG. That is, while a single participant could theoretically mention a particular category of social-cognitive reasoning to explain their partners actions up to three times during the dissolution phase, they still only received a frequency count of one. Therefore, frequency scores meet the assumption of independence for each phase. Due to the modest frequency counts for some categories, the results will highlight the most frequently endorsed categories by each group for each phase of the game. Furthermore, to avoid reporting on negligible differences only differences in proportion of at least 25% or greater between the high and low BPD groups will be described in the results. In terms of number of participants, this reflects a difference of at least four or more participants. As we aimed to understand the social-cognitive reasoning behind trust patterns observed in our previous research, the results will highlight group differences in social-cognitive reasoning during the (a) overall trust game; (b) formation phase; (c) dissolution phase; and (d) the restoration phase *relative* to the formation phase.

Table 12*Social-Cognitive Reasoning for Self and Partner Trust Game Behaviours as Trust Forms, Dissolves, and Restores*

	Formation Phase				Dissolution Phase				Restoration Phase				All Phases			
	Self		Partner		Self		Partner		Self		Partner		Self		Partner	
Social-Cognitive Reasoning Category	High BPD <i>n</i> (%)	Low BPD <i>n</i> (%)	High BPD <i>n</i> (%)	Low BPD <i>n</i> (%)	High BPD <i>n</i> (%)	Low BPD <i>n</i> (%)	High BPD <i>n</i> (%)	Low BPD <i>n</i> (%)	High BPD <i>n</i> (%)	Low BPD <i>n</i> (%)	High BPD <i>n</i> (%)	Low BPD <i>n</i> (%)	High BPD <i>n</i> (%)	Low BPD <i>n</i> (%)	High BPD <i>n</i> (%)	Low BPD <i>n</i> (%)
Fairness	1 (6.3)	2 (12.5)	- (18.8)	3 (18.8)	- (18.8)	3 (18.8)	- (18.8)	1 (6.3)	1 (6.3)	2 (12.5)	1 (6.3)	4 (25.0)	3 (18.8)	4 (25.0)	1 (6.3)	5 (31.3)
Prosocial	4 (25.0)	4 (25.0)	2 (12.5)	- (12.5)	2 (12.5)	2 (12.5)	1 (6.3)	- (12.5)	- (12.5)	4 (25.0)	- (12.5)	2 (12.5)	5 (31.3)	6 (37.5)	3 (18.8)	2 (12.5)
Positive reciprocity	1 (6.3)	4 (25.0)	7 (43.8)	4 (25.0)	- (12.5)	- (12.5)	- (12.5)	1 (6.3)	1 (6.3)	2 (12.5)	7 (43.8)	8 (50.0)	1 (6.3)	5 (31.3)	10 (62.5)	10 (62.5)
Trust	10 (62.5)	10 (62.5)	4 (25.0)	4 (25.0)	7 (43.8)	5 (31.3)	2 (12.5)	2 (12.5)	9 (56.3)	6 (37.5)	7 (43.8)	7 (43.8)	12 (75.0)	12 (75.0)	7 (43.8)	9 (56.3)
Mistrust	2 (12.5)	5 (31.3)	4 (25.0)	3 (18.8)	2 (12.5)	2 (12.5)	4 (25.0)	4 (25.0)	1 (6.3)	3 (18.8)	- (12.5)	4 (25.0)	4 (25.0)	7 (43.8)	6 (37.5)	7 (43.8)
Negative reciprocity	6 (37.5)	2 (12.5)	2 (12.5)	1 (6.3)	9 (56.3)	9 (56.3)	7 (43.8)	5 (31.3)	1 (6.3)	1 (6.3)	2 (12.5)	2 (12.5)	11 (68.8)	10 (62.5)	8 (50.0)	7 (43.8)
Personal gain	2 (12.5)	- (12.5)	5 (31.3)	6 (37.5)	4 (25.0)	5 (31.3)	4 (25.0)	7 (43.8)	1 (6.3)	4 (25.0)	2 (12.5)	3 (18.8)	5 (31.3)	6 (37.5)	7 (43.8)	8 (50.0)
Negative partner traits	1 (6.3)	1 (6.3)	6 (37.5)	2 (12.5)	- (12.5)	1 (6.3)	6 (37.5)	6 (37.5)	- (12.5)	- (12.5)	2 (12.5)	1 (6.3)	1 (6.3)	2 (12.5)	10 (62.5)	7 (43.8)
Negative affect	1 (6.3)	2 (12.5)	- (12.5)	1 (6.3)	2 (12.5)	3 (18.8)	1 (6.3)	4 (25.0)	1 (6.3)	3 (18.8)	- (12.5)	2 (12.5)	4 (25.0)	5 (31.3)	1 (6.3)	5 (31.3)
'Don't know'	5 (31.3)	1 (6.3)	2 (12.5)	1 (6.3)	3 (18.8)	1 (6.3)	6 (37.5)	3 (18.8)	5 (31.3)	2 (12.5)	8 (50.0)	4 (25.0)	7 (43.8)	2 (12.5)	11 (68.8)	5 (31.3)
Other	6 (37.5)	6 (37.5)	2 (12.5)	5 (31.3)	3 (18.8)	6 (37.5)	4 (25.0)	6 (37.5)	5 (31.3)	8 (50.0)	5 (31.3)	9 (56.3)	8 (50.0)	11 (68.8)	6 (37.5)	12 (75.0)
Total	39	37	34	30	32	37	35	39	25	35	34	46	61	70	70	77

Social-cognitive reasoning about self and partner across all phases

Across the TG the high BPD group compared to the low BPD group were more likely to respond that they did not know when attempting to explain their own behaviour (43.8% vs 12.5%) or their partners behaviour (68.8% vs 31.3%). The high BPD group were less likely than the low BPD group to attribute their own behaviour to positive reciprocity (6.3% vs 31.3%). The high BPD group were also less likely than the low BPD group to attribute partner behaviour to fairness (6.3% vs 31.3%) and negative affect (6.3% vs 31.3%).

Social-cognitive reasoning in the trust formation phase

Reasoning about self

During the formation phase, trust was the most frequently made self-attribution by both groups. Approximately three times as many participants in the high BPD group were more likely to use negative reciprocity to explain their own behaviour (37.5% vs 12.5%). Notably, almost one third of the high BPD group mentioned that they did not know why they behaved as they did during this phase (31.3%), which was markedly lower in the low BPD group (6.3%).

Reasoning about partner

During the formation phase, the most frequently made partner-attribution by the high BPD group was positive reciprocity, while the most frequently made partner-attribution by the low BPD group was personal gain. During the formation phase, 37.5% of the high BPD group used negative partner traits to explain their partner's behaviour compared to 12.5% of the low BPD group.

Social-cognitive reasoning in the trust dissolution phase

Reasoning about self

During the dissolution phase, the most frequent self-attributions made by both groups was negative reciprocity. There were no differences of at least 25% between the groups in self-attributions made during the dissolution phase.

Reasoning about partner

During the dissolution phase, the most frequently made partner-attribution by the high BPD group was negative reciprocity, while the most frequently made partner-attribution by the low BPD group was personal gain. There were no differences of at least 25% between the groups in partner-attributions made during the dissolution phase.

Social-cognitive reasoning in the trust restoration phase relative to social-cognitive reasoning in the trust formation phase

Reasoning about self

While 37.5% of the high BPD group had used negative reciprocity to describe their own behaviour during the trust formation phase, only 6.3% used negative reciprocity self-attributions during the trust restoration phase. Similarly, while 25% of the high BPD group had made prosocial attributions to describe their own behaviour during the trust formation phase, none used prosocial self-attributions during the trust restoration phase. For the low BPD group, the proportion using trust-based reasoning to explain their own behaviour fell from 62.5% during the formation phase to 37.5% during the restoration phase, while personal gain self-attributions increased from 0% to 25.0% during the same period.

Reasoning about partner

During the restoration phase, the most frequent partner-attribution made by the high BPD group was ‘don’t know’ (50.0%), which was only made by 12.5% of the high BPD group during the formation phase. While 37.5% of the high BPD group had used negative partner traits to explain their partner’s behaviour during the trust formation phase, only 12.5% used it to do so during the trust restoration phase. During the restoration phase, 50.0% of the low BPD group made positive reciprocity attributions to explain their partner’s behaviour, which was double that did during the formation phase.

Discussion

In recent decades, the maladaptive interpersonal functioning indicative of BPD has been examined through the lens of anomalies in social cognition (for reviews see Dinsdale & Crespi, 2013; Lazarus et al., 2014; Németh et al., 2018; Roepke et al., 2012; Sharp, 2014). Research suggests BPD is associated with deficits in reasoning about the mental states of self and others (see Bora, 2021), including the tendency to hypermentalize by making inferences about another person's mental states that far exceed what could be justified by the observable data (Andreou et al., 2015; Cortés-García et al., 2021; Penner et al., 2020; Quek et al., 2019; Sharp et al., 2013; Sharp, Pane, et al., 2011; Somma et al., 2019; Vaskinn et al., 2015). The current study aimed to examine the social-cognitive reasoning associated with trust behaviours in a TG for individuals with a high BPD trait count versus individuals with a low BPD trait count. The reasons provided fell into nine thematic categories of reasoning: fairness, prosocial, positive reciprocity, trust, mistrust, negative reciprocity, personal gain, trait, and negative affect, as well as ‘don’t know’.

The analysis is predominantly descriptive and aimed to clarify why BPD was associated with increased mistrust during initial interactions (formation phase) with a partner (Unoka et al., 2009), even when the partner was being cooperative (Abramov et al., 2020; Liebke et al., 2018). Previous research has found a negative trust bias associated with BPD (for review see Fertuck et al., 2018; Masland et al., 2020), characterised by views of others as deceitful, exploitative, and betraying (Barazandeh et al., 2016; Bhar et al., 2008), and perceiving others as rejecting (De Panfilis et al., 2015; Domsalla et al., 2014; Renneberg et al., 2012; Staebler et al., 2011; Weinbrecht et al., 2018). Based on these findings it was assumed that high BPD trait participants approached the TG expecting the worst. This was supported by our previous findings that when baseline feelings of rejection and beliefs about needing to protect oneself were taken into account, BPD trait count was no longer associated with increased mistrust during the initial stage of a TG with a cooperative partner (Abramov et al., 2022).

Results from the current study revealed that the high BPD group were more likely to apply negative characteristics to their partner such as “cheap” or “rude” to explain the partner’s behaviour during the formation phase, whereas the low BPD group were more likely to infer that their partner was motivated by personal gain. While both types of reasoning paint a picture of the partner as being self-serving, the focus on negative partner traits may imply greater hostility, both from the partner, but also towards the partner. This is consistent with the negative beliefs of others (Barazandeh et al., 2016; Bhar et al., 2008) and the hostile-interpretation bias that has been associated with BPD (Arntz et al., 2011; Barnow et al., 2009; Lobbestael & McNally, 2016).

To explain their own behaviour during the trust formation phase, the high BPD group were more likely to reason using negative reciprocity, yet the high BPD group also reasoned that their partner was motivated by positive reciprocity. The latter suggests that the high BPD group were able to recognize that their partner was returning a profit, which is consistent with how the game was programmed, and also consistent with previous findings that individuals with BPD do not differ from controls in their accuracy in recognizing the fairness of investments in a TG (Franzen et al., 2011). However, in explaining their own behaviour as a tit-for-tat strategy, and using derogatory terms to describe their partner, it is possible that despite recognizing that their partner was returning a profit, the high BPD group nevertheless responded in a punitive manner. This was quantitatively reflected in the slower rate of trust growth associated with BPD trait count during the formation phase (Abramov et al., 2020).

This is consistent with existing research that prosocial behaviour may evoke a negative response in those with borderline pathology. Research suggests that individuals with BPD features respond to prosocial behaviour with less positive emotions compared to controls (Bhatia et al., 2013; Reichenberger et al., 2017; Sadikaj et al., 2010). Individuals with BPD may even penalize positive behaviour directed at themselves. For example, Liebke and colleagues (2018) found that after receiving social acceptance feedback, the BPD group were less cooperative in a subsequent but unrelated social encounter, sending less money in a TG than controls. In fact, BPD patients appeared to be less trusting after being socially accepted rather than rejected, and this became more pronounced the lower the expectation of acceptance had been (Liebke et al., 2018). This suggests that the increased dissonance between an expectation of rejection and actual social acceptance may have elicited greater general mistrust or suspicion regarding the intentions of the other. In another study, compared to HCs, individuals with BPD exhibit a more punishing response to fair offers in an economic

game by rejecting fair offers more frequently than healthy control's, while BPD status was not found to impact response to unfair offer (De Panfilis et al., 2019). An expectation of betrayal may not only override experiences to the contrary, such as a cooperative partner, but potentially trigger a 'strike first' stance in the borderline personality.

This is consistent with findings linking hostility reactivity to perceived rejection in individuals with borderline personality pathology (Hepp, Lane, et al., 2018; Richmond et al., 2020; Scott et al., 2017).

The current study also aimed to use cognitive reasoning to explain the paradoxically increasing trust associated with BPD during the dissolution phase (Abramov et al., 2020). Comparison of reasoning between groups highlighted both similarities and differences. The groups did not differ to a marked extent in how likely they were to use trust or mistrust reasoning to explain their own or their partner's behaviour. During the dissolution phase the most frequently reported category of reasoning by both groups was negative reciprocity to explain their own behaviour. However, previous research found that BPD trait count was associated with entrusting progressively larger sums of money in response to trust violating partner behaviour (Abramov et al., 2020). The incongruence between the meaning ascribed to their own behaviour, a tit-for-tat strategy, and their actual behaviour, to progressively send their partner more money, may indicate difficulties with interpreting their own behaviour, which could reflect potential mentalization difficulties. Difficulty providing reasons for their own and their partners behaviour was more prevalent throughout each phase of the game for the high BPD group, although it did not reach the 25% or greater difference threshold during the dissolution phase. These findings are consistent with the broader literature on mentalization deficits in BPD (Bora, 2021).

During the trust dissolution phase the low BPD group were also more likely to reason that their partner's actions were related to a desire for personal gain, thereby continuing that reasoning from the formation phase. In contrast, the high BPD group were now more likely to describe their partner as engaging in a tit-for-tat strategy, while during the trust formation phase they had reasoned that their partner was engaging in positive reciprocity. Given that the high BPD group also described their own behaviour during the trust formation phase as a tit-for-tat response, it suggests they may have interpreted the violations by their partner as a punitive response to their own punitive behaviour, rather than a desire for personal gain as had been reasoned by the low BPD group. Viewed this way it is possible that the increased trust behaviours observed in our previous research (Abramov et al., 2020), reflect an attempt to make amends based on an interpretation of the trust violations as a punitive response to their own previously hostile behaviour. While this would be in contrast to King-Casas et al.'s (2008) findings that BPD trustees did not engage in coaxing behaviours, the difference may be a function of the different roles the participants played. In their study, participants played the role of trustee whereas in the current study, they play the role of trustor which is an inherently more powerful role given the trustee is reliant on the trustor to invest in the first place.

Finally, our previous research found that for high BPD trait individuals, trust restored faster than it was formed (Abramov et al., 2020). As such if we compare the reasoning made during restoration to that made during formation, we see that for the high BPD group, negative reciprocity was mentioned by approximately six times as many individuals during formation than during restoration to explain their own behaviour. This suggests a less punitive style of investing behaviour during restoration than formation for the high BPD group. In other words, the reasoning they provide to explain their behaviour is congruent with

their behaviour that they are more willing to engage in trusting behaviours following a violation than prior to any violation when trust is forming.

Similarly, when reasoning about their partner's behaviour, there were fewer high BPD participants referring to negative partner traits in the restoration phase whereas this had been one of their most frequent partner attributions during the formation phase. Moreover, twice as many of the high BPD group compared to the low BPD group stated that they did not know why their partner behaved as they did during the restoration phase, which was a four-fold increase from the formation phase. These findings suggest that the high BPD group were less combative and perhaps more confused, when their partner resumed cooperative play after multiple trust violations, than at the beginning of the game when their partner had been cooperative. As suggested earlier, it is possible that individuals with high BPD traits begin interactions with new partners expecting the worst and therefore behaving in a more self-protective and readily punitive manner. The current findings suggest that individuals with a borderline personality structure may recognize this defensive style of behaviour, and at some level, accept responsibility for it. Hence there may be in part a sense of relief when the betrayal occurs, but also a sense that they need to make amends.

What does this mean for people with high levels of BPD traits? First, it is likely that holding negative core beliefs about others and ones need to protect oneself compromises the ability of these individuals to reason in a socially normative fashion. It also potentially compromises their ability to create cooperative, harmonious interactions with new people. Even when they interact with partners who they recognize as objectively cooperative, high BPD trait individuals may behave in hostile ways. This hostility however appears to stop and even reverse when the partner retaliates, and the high BPD trait individual appears to

recognize that this retaliation was prompted by their own earlier hostility. Previous research has already shown that people with BPD may make negative first impressions on others, being rated as less likeable, trustworthy, and cooperative than controls based on short film clips where they described their personal preferences such as hobbies (Hepp, Störkel, et al., 2018). The findings of the current study capture the negative, hostile social-cognitive reasoning processes that potentially underlie mistrustful behaviours associated with BPD trait count when engaging with a new and cooperative partner (Abramov et al., 2020). Our collection of studies based on quantifying trust within a social exchange offers a glimpse into those characteristics, behaviours, and cognitive reasoning processes that may render individuals with BPD more likely to make a negative first impression, and perhaps result in missed opportunities for the development of new connections. Additionally, the paradoxically less hostile reasoning reported by the high BPD group after the trust violation compared to the period when trust should ostensibly be forming, and their more trusting behaviours after the violation (Abramov et al., 2020), reflects a dynamic that would likely be confusing to many people, and perhaps hinder the development of stable relationships.

Finally, given that the high BPD group were consistently less able to explain their own or their partners behaviour, and this increased as the game progressed, therapeutic interventions that encourage and develop social-cognitive reasoning about self and others may be warranted. Mentalization-based therapy purports to increase mentalization skills in order to improve interpersonal functioning, and there are promising findings suggesting its effect on improving social and interpersonal functioning (for review see Malda-Castillo et al., 2018). For example, mentalization-based therapy has been found to result in greater reductions in interpersonal problems, especially mistrust, for patients with BPD compared to a long-term specialized psychodynamic treatment program (Kvarstein et al., 2015).

There are a number of limitations to the current study. First, participants were asked to provide reasons for their own and their partner's behaviour at the end of the TG. Although the protocol reminded participants how much had been sent and returned for the relevant rounds, the reasoning was nevertheless conducted several minutes after the transaction and may therefore have not captured the automatic thought processes that occurred during the TG itself. Future research could elicit reasoning at the time of investment although this may promote explicit mentalization that may influence subsequent TG transactions. Second, only an open response format for social-cognitive reasoning was used which then had to be coded into categories. To increase reliability, future research could also trial pre-set responses (e.g., as generated by this study). Third, with only 16 participants in the BPD group the sample size was modest. To account for this, reporting of between-group differences was limited to instances where there was at least a 25% or greater difference in the proportion of each group that endorsed the reasoning category. Fourth, a university student sample was used, and groups were based on scores on a self-report measure of BPD traits as opposed to use of a clinically diagnosed sample of people with BPD. Although significant effects for differences in the TG were found with such a sample (Abramov et al., 2020, 2022), the research should be replicated with a clinical sample.

Given that relationship instability is a key feature of borderline pathology, understanding the reasoning behind both the behaviours that threaten relationships, such as lack of trust and cooperation in new social encounters, and the behaviours that maintain relationships, even in the face of ruptures, is essential to helping this population improve interpersonal dynamics. The current study was the first in the BPD literature to examine the social-cognitive reasoning behind behaviour in a trust-based social exchange. Based on our

findings areas for potential intervention include addressing the negative and hostile beliefs about others, and expectations of betrayal that individuals with BPD may hold. In addition, these individuals may benefit from training to improve social-cognitive reasoning skills, in particular to help them understand how incongruent behaviours, such as responding to positive reciprocity with a tit-for-tat strategy, may be experienced by others.

Chapter 5: General Discussion

Summary of Findings

Interpersonal dysfunction is a defining characteristic of BPD (Gunderson, 2007), and trust is proposed to be one of the core underpinnings of this relational disturbance (Masland et al., 2020; Poggi et al., 2019). BPD is associated with negative core beliefs regarding the intentions of others including expectations of betrayal, abuse, and rejection (Barazandeh et al., 2016; Bhar et al., 2008), and endorsing the view that one should pre-emptively act to protect oneself (Bhar et al., 2008). Individuals with BPD tend to appraise others as untrustworthy (Fertuck et al., 2013; Masland & Hooley, 2020; Miano, Fertuck, et al., 2017), and exhibit less cooperative behaviours in trust-based interpersonal exchanges (King-Casas et al., 2008; Liebke et al., 2018; Roberts et al., 2018; Unoka et al., 2009). Paradoxically, trustworthiness or actual cooperation by another appears to compound the mistrustful behaviour associated with BPD (Liebke et al., 2018; Niedtfeld & Kroneisen, 2020).

The overarching aim of this thesis was to further understand the anomalous interpersonal trust behaviours for people with traits of BPD. There is insufficient use of research design and analytical methods that adequately capture the dynamic nature of trust (Korsgaard et al., 2018; Lewicki et al., 2006; Rousseau et al., 1998), as well as a lack of empirical inquiry into the mechanisms underpinning the trust behaviours associated with BPD. These gaps in the knowledge were the impetus for the development of the three studies that comprise this thesis. Study 1 set out to examine whether the number of BPD traits endorsed by the individual influenced the trajectory of trust as it formed with a new partner, dissolved in response to trust violation, and restored in response to trust repair. Study 2 aimed to build on this by determining whether the effect of BPD trait count on these trust patterns

was moderated by attachment style, self-protective beliefs, or feelings of rejection. Study 3 sought to examine how the social-cognitive reasoning given to explain the TG behaviour of oneself or one's partner varied according to whether a high or low BPD trait count was endorsed.

As presented in Chapter 2, Study 1 combined an iterative TG protocol with a data analytical technique suited to modelling data in which discontinuities are present to capture the direction, speed, and magnitude of changes in trust behaviours as a function of BPD trait count. Trustee reciprocity was programmed to create a trust violation to elicit the dissolution of trust, and a trust repair bid to elicit the restoration of trust. DGM revealed the trajectories of trust as it formed, dissolved, and restored both across individuals and within individuals, the latter measuring how trust changed in response to violation and repair events relative to the level and growth rate of trust prior to these events occurring. It was found that BPD trait count was associated with a paradoxical pattern of trust behaviour. Specifically, during the initial phase of interacting with a new and cooperative partner, BPD trait count was associated with declining trust, but in response to multiple consecutive instances of trust violation by that partner, BPD trait count was unexpectedly associated with increasing trust. Moreover, by utilizing a coding structure to capture the relative intraindividual fluctuations in trust (Bliese & Lang, 2016), BPD trait count was positively associated with trust restoring at a relatively faster rate than it was formed. Finally, BPD trait count was not associated with post-game trustworthiness or fairness partner appraisals.

In order to understand what factors might explain the effect of BPD on trust dynamics, a number of constructs conceptually and/or empirically related to interpersonal difficulties in BPD and trust were selected for observation in Study 2. Building on the same

model as derived in Study 1, Study 2 added variables of interest to determine whether they accounted for or moderated the effect of BPD trait count on trust behaviours. Findings revealed that pre-existing feelings of rejection and self-protective beliefs each predicted a slower rate of trust formation when BPD trait count was accounted for, with trait count no longer a significant predictor of trust growth during this phase. Additionally, there was a trend towards self-protective beliefs moderating the effect of BPD on trust formation, by intensifying the trust negating effect of BPD trait count during this stage. In contrast, endorsing a preoccupied attachment style positively moderated the effect of BPD trait count when trust was forming and immediately after the first repair bid, in other words, having a trust-enhancing effect. Study 2 also revealed that once self-protective beliefs were accounted for, the paradoxically faster rate of trust growth associated with BPD during the dissolution phase no longer met significance. While self-protective beliefs failed to explain a significant proportion of the variance observed, the findings suggest that the faster rate of trust growth during the dissolution phase observed in those with a high number of BPD traits may reflect some attempt at self-preservation.

The final step in this body of work, as represented by Study 3, was to determine whether there were discernible differences according to BPD trait count in how individuals reasoned during a TG when contemplating their partners' and their own behaviour. Nine categories of reasoning were identified: fairness, prosocial, positive reciprocity, trust, mistrust, negative reciprocity, personal gain, negative character trait, and negative affect, as well as a category that reflected not knowing ('don't know'). Findings revealed that during the trust formation phase, the high BPD trait count group appeared more likely than the low BPD trait count group to explain their own behaviour as negative reciprocity (tit-for-tat strategy) and to explain their partner's behaviour by attributing negative characteristic traits

(e.g., “rude”, “cheap”, “asshole”). Yet, the most frequent inference made about partner behaviour by the high BPD group during this phase was positive reciprocity. Furthermore, the high BPD group used negative reciprocity to describe their own behaviour and negative character traits to describe their partner’s behaviour markedly *less* often during the restoration phase compared to the formation phase. That is, they appeared to exhibit more hostility before the trust violation had occurred and paradoxically, less after. Study 3 failed to find any specific categories of reasoning that differed markedly between the high and low BPD trait count groups during the trust dissolution phase. Finally, there appeared to be a general tendency across the combined three phases of the TG for high BPD trait count individuals to state that they were unsure or did not know why they or their partner behaved as they did.

Implications

Empirical and Conceptual Implications

Study 1 revealed that BPD trait count was negatively associated with trust formation, with high BPD trait count participants sending progressively smaller amounts to the trustee across the first five rounds of the TG, which was consistent with the reasoning of negative reciprocity the high BPD trait count group used to explain their own behaviour in Study 3. Study 2 indicated that the trust-suppressing effects of BPD trait count during the formation phase were accounted for by pre-existing feelings of rejection and self-protective beliefs, and Study 3 revealed that the high BPD trait group were more likely than their low BPD counterparts to attribute negative characteristics to the trustee during this early phase of the game. Importantly, the most common reasoning category used by high BPD trait count participants to describe trustee responses during the trust formation phase was positive

reciprocity, which was consistent with how the TG was programmed. In other words, despite an objective and subjective experience of a cooperative partner, BPD trait count was associated with greater hostility in their own reasoning and when making partner attributions, and increasingly *less* trusting behaviour, with self-protective beliefs and pre-existing feelings of rejection appearing to explain the mistrustful behaviours.

These findings corroborate earlier findings that people with BPD are more mistrustful than HCs in their beliefs (Barazandeh et al., 2016; Barnow et al., 2009; Bhar et al., 2008; Segal et al., 1992, 1993) and behaviour (Liebke et al., 2018; Niedtfeld & Kroneisen, 2020; Roberts et al., 2018; Unoka et al., 2009). Compared to HCs, people with BPD are less trusting when playing a TG with a new partner (Unoka et al., 2009), even when the other player is being cooperative (Liebke et al., 2018), are more likely to expect others to behave selfishly (Lévay et al., 2021), and to hold pessimistic post-game predictions about the outcome of a TG even when there is no evidence in either direction (Unoka et al., 2009). The current research compliments the existing literature by showing that endorsing self-protective beliefs accounts for the diminishing trust behaviours associated with BPD features when interacting with a new and cooperative partner. Interestingly, while much of the research suggests that relative to HCs, people with BPD are more likely to make negatively biased trust appraisals (Fertuck et al., 2013; Masland & Hooley, 2020; Miano et al., 2013; Nicol et al., 2013; Richetin et al., 2018), this was not supported by the post-game appraisals of partner trustworthiness or fairness observed in Study 1. Additionally, Study 3 showed that the high BPD trait count group recognized positive reciprocity in the other players behaviour. In other words, BPD trait count does not appear to interfere with recognizing objectively cooperative behaviour (Franzen et al., 2011) which suggests that the untrusting and/or hostile behaviour observed in high BPD trait count individuals is likely to be influenced considerably by global

beliefs about people in general rather than specific to the other player. That is, it is the sweeping stories that individuals with BPD hold about trust, betrayal, their own vulnerabilities, and how to protect themselves, that influence their trust behaviours when forming trust with a new party.

Ebert et al. (2013) suggested that due to experiences of childhood adversity, individuals with BPD may expect rejection or abandonment and therefore not be motivated to cooperate. It is important at this point to distinguish between socially normative lapses in cooperation and uncooperative acts which may be underpinned by borderline vulnerabilities. King-Casas et al. (2008) demonstrated that *both* healthy and borderline trustees appeared to behave in self-serving ways over time by gradually attempting to keep more of the funds for themselves. Cochara et al. (2004, p. 41) observed this phenomenon in non-clinical populations and referred to it as “erosion of reciprocity” suggesting that in iterative investment games, initially there is an impetus to instil confidence in the other player that one is trustworthy, and that once this ‘reputation’ has been established, individuals may adopt more self-interested behaviours. Faced with erosion of reciprocity, the trustors in King-Casas et al. (2008) retaliated with smaller investments to which healthy trustees responded with coaxing behaviours in the form of higher returns, while their BPD counterparts further reduced the amount returned or withheld altogether, and were more likely to experience total breakdown in cooperation. That King-Casas et al.’s (2008) BPD trustees chose to respond in ways that further jeopardized trust rather than repaired it may reflect a resigned response to a worldview that others will eventually betray, so it is preferable to strike first, rather than interpreting the drop in investment size as a prompt to remedy eroding reciprocity. Individuals high in BPD traits may begin trust-based exchanges waiting for the other shoe to drop, and therefore pre-emptively adopt a retaliatory, self-protective strategy. In accordance

with this, experimental research using other game paradigms have found that compared to HCs, individuals with BPD have been found to engage in more behavioural aggression (Dougherty et al., 1999; McCloskey et al., 2009; New et al., 2009) and behave less cooperatively (Saunders et al., 2015).

The role of rejection-related affect may also need to be considered. Poggi and colleagues (2019) proposed that in addition to negatively-biased trust appraisals, rejection sensitivity may play a mechanistic role in interpersonal disturbances observed in BPD. Borderline pathology is associated with a greater likelihood of interpreting the behaviour of others as rejecting (Lobbestael & McNally, 2016) and hostile (Arntz et al., 2011; Barnow et al., 2009; Lobbestael & McNally, 2016), and individuals with BPD are more likely to feel that they have been excluded even when objectively, this is not the case (De Panfilis et al., 2015; Domsalla et al., 2014; Renneberg et al., 2012; Staebler et al., 2011; Weinbrecht et al., 2018). Both the anxious expectation that one will be rejected (Ayduk et al., 2008), and perceived or experienced rejection (Berenson et al., 2011; Hepp, Lane, et al., 2018; Lazarus et al., 2018; Richmond et al., 2020; Scott et al., 2017) are associated with increased hostility for individuals with BPD. In a similar vein, Study 2 revealed that feelings of rejection that were present before the TG were also associated with declining trust during the formation phase and accounted for the trust-diminishing effect of BPD trait count. Interestingly, acetaminophen (i.e., paracetamol), which has been shown to reduce self-reported social pain (hurt feelings) from social rejection (DeWall et al., 2010; Fung & Alden, 2017), is associated with a reduction in mistrustful TG behaviours in individuals with high levels of BPD features (Roberts et al., 2018). It is possible that social exchanges in general elicit expectations of maltreatment by others, with betrayal and rejection being so intimately entwined that we cannot be sure that for individuals with high levels of BPD features, the TG is capturing

issues with trusting alone or pre-emptive efforts to avoid a more general maltreatment by others which may include acts of betrayal, rejection, abandonment, etc.

Expectations of betrayal or general maltreatment may also mean that people with a higher BPD trait count are hypersensitive to cues signalling social threat (Bertsch et al., 2018). In the current thesis small fluctuations in reciprocity were programmed into the TG within each phase. For example, during the trust formation phase trustee repayments for each round were randomized to fall between 34% to 44% of the tripled investment. As such, investing 50 MU for each of the first five rounds could theoretically net five different returns fluctuating between 51 and 66 MU. It is possible that higher BPD trait count individuals were more sensitive to the presence of fluctuating returns, and perhaps interpreted the variability in returns as threatening or even confusing, encouraging a self-protective or hostile stance. The ability to ‘read’ the fluctuations as either benign, a ‘testing of the limits’, an erosion of reciprocity, or signs of impending betrayal may hinge on mentalization capabilities. The literature suggests that the mentalization difficulties observed in individuals with BPD are apparent with more emotionally charged, complex or ambiguous social-cognitive tasks (Daros et al., 2013; Minzenberg et al., 2006; Mitchell et al., 2014; Richman & Unoka, 2015; Sharp & Vanwoerden, 2015). Study 3 revealed that across the TG in general, the high BPD trait count group more frequently than the low BPD trait count group were unable to provide a reason for their own or their partners behaviour, and at times there was an incongruence between their reasoning and their behaviour. It is possible these anomalies reflect mentalization deficits.

It is also important to consider what effects uncertainty or ‘not knowing’ may have on an individual’s sense of competency to navigate ambiguous social situations. Individuals low

in BPD traits may view a degree of power struggle or ‘erosion of reciprocity’ as normal and expected, and superior mentalizing capacities may enable them to distinguish, decipher, and respond to meaningful fluctuations more effectively without compromising the relationship, such as by adopting coaxing-style behaviours to reinforce greater cooperation (Barton et al., 2021; King-Casas et al., 2008). In contrast, impaired mentalization, in particular uncertainty about the mental state of others, may impact the individual’s sense of confidence that they can understand and respond effectively to signals such as the fluctuating trustee returns. There is mixed evidence regarding the confidence that individuals with BPD express in reference to their appraisal judgments regarding emotion in others with support for both a lack of confidence and overconfidence (Kaletsch et al., 2014; Schilling et al., 2012; Thome et al., 2016). In Study 3, in addition to being less able to provide reasons for the other players decisions, the high BPD group were also more likely to attribute negative characteristics and behave in a mistrustful manner to an otherwise cooperative partner during the initial phase of the game. These types of attributions and hostile behaviour may be related to hypermentalization, that is, an overattribution of mental states that exceed the observable evidence (Sharp et al., 2013). To conclude that the other players behaviour is due to their being “cheap” or “rude” may also reflect overcompensatory efforts to cope with the lack of comprehension (reasoning) behind both their own and the other players behaviour which may be potentially experienced in more nuanced or ambiguous interpersonal contexts, such as, the subtly fluctuating returns in a TG. In other words, the high BPD trait count group may have been experiencing both diminished mentalization as reflected in the increased reports of ‘not knowing’ and hypermentalization.

In more recent years, Fonagy and colleagues have expanded upon the mentalization-based theory of BPD to viewing the disorder as a failure to develop sufficient epistemic trust

(Fonagy et al., 2015). Epistemic trust refers to the perception that information conveyed interpersonally is true, reliable, and personally relevant, allowing the individual to learn from their social environment and to integrate these learnings into their existing repository of culturally-transmitted knowledge (Fonagy et al., 2015; Sperber et al., 2010). Until recently (C. Campbell et al., 2021), the lack of a validated instrument to assess epistemic trust has meant its association with BPD has remained largely theoretical. However, there is emerging empirical support for the relationship between deficits in epistemic trust and BPD (Orme et al., 2019), as well as theoretically-congruent findings such as the observation that BPD is associated with a failure to learn from positive social feedback (Liebke et al., 2018). Although the examination of epistemic trust was beyond the scope of this thesis, the progressively smaller investments sent during the trust formation phase despite acknowledging the trustee was reciprocating positively may indicate a failure to incorporate and adapt to this prosocial information. Moreover, the more frequent lack of reasoning or uncertainty expressed to explain partner behaviour observed in the high BPD trait count group may suggest a failure to see the social cues as true, reliable, and/or personally relevant, potentially encouraging a stance informed by generalised core mistrustful beliefs rather than in-the-moment social feedback.

An alternative explanation for the lack of cooperation associated with BPD is proposed to stem from the holding of aberrant social norms, where positive social cues are responded to unfavourably (King-Casas et al., 2008; Liebke et al., 2018). Compared to HCs, BPD has been associated with less positive emotions (Bhatia et al., 2013; Reichenberger et al., 2017; Sadikaj et al., 2010) and a greater likelihood of hostility (De Panfilis et al., 2019) in response to prosocial behaviour. Liebke et al. (2018) found that following priming with social acceptance, trustors with BPD engaged in less trusting behaviours with cooperative trustees

than after being primed with social rejection. This implies positive social cues, such as acceptance, appear to heighten mistrust for people with BPD. Similarly, using an EEG-compatible version of Cyberball (Williams & Jarvis, 2006), under the social inclusion condition, individuals with BPD in comparison to HCs were found to have an increased P3 amplitude, which is an EEG indicator for expectancy violation (Gutz et al., 2015; Weinbrecht et al., 2018). Furthermore, they also reported more negative mood, ostracism, and threat to social needs when socially included than reported by HCs. Positive social cues may further violate the expectation of maltreatment that individuals with BPD have and potentially trigger a self-protective, hostile stance. Interestingly, Liebke et al (2018) also found that while HCs and individuals with BPD adjusted their expectations of future rejection after experiencing rejection, only the individuals with BPD failed to adjust their expectation of future acceptance after experiencing acceptance. The authors questioned whether individuals with BPD deviate from the norm in how they appraise and integrate *positive* social cues, resulting in a failure to respond in kind. As proposed earlier, this may reflect failures in epistemic trust.

If the trust formation phase captured the more hostile less trusting side of BPD pathology, the trust dissolution phase turned this behaviour on its head. It is at this point that the methodological advantages of Study 1 are realised since if only the formation phase were considered, then it might erroneously be assumed that BPD trait count is uniformly associated with higher levels of mistrust and unwillingness to cooperate in trust-based exchanges. However as seen in Study 1, while the trust violation was occurring, BPD trait count was associated with making progressively larger investments despite sustaining repeated violations. Further, the rate at which trust was growing slowed down *after* the programmed trustee sent back a markedly larger sum signalling a bid to repair trust. In fact,

this ‘slowing down’ was moderated by BPD trait count so that those with the highest numbers of symptoms paradoxically responded with a drop in investment size to the first repair bid. Moreover, as BPD trait count increased, trust restored at a faster rate than it was formed, which was congruent with the reasoning provided in Study 3, in which the high BPD trait count group were more likely to use negative reciprocity to describe their own behaviour during the formation phase than during the restoration phase.

As proposed earlier, it is possible that the paradoxical trust behaviours associated with BPD trait count during and after the trust violation could be attributed to aberrations in social norms. There is evidence based on other economic game paradigms that BPD is associated with accepting and perhaps even opting for inequitable treatment (De Panfilis et al., 2019; Jeung et al., 2020; Polgar et al., 2014). Compared to HC’s, individuals with BPD appear to be less discriminatory in their approach/avoidance responses to negative social stimuli (Schneider et al., 2020). In a Cyberball experiment examining approach-avoidance behaviour to positive or negative emotional stimuli (facial expressions) after experiencing social exclusion, patients with BPD were found to approach happy faces less than HC’s. They were also less discriminating in their approach-avoidance response to happy or angry faces, in comparison to HC’s who ‘approached’ happy faces more than angry faces (Wiesenfeller et al., 2020). BPD features have also been associated with increased intimate partner or dating victimization, and higher risk of future victimization (Cavelti et al., 2021; Maneta et al., 2013; Reuter et al., 2015; Vanwoerden et al., 2019), which may reflect a greater tolerance of negative interpersonal dynamics. On the other hand, the violation (the ‘other shoe dropping’), may have come as a relief to those high in borderline traits, affirming that their suspicions had been founded, perhaps even giving them the confidence to now engage more fully knowing the scope of behaviour (betrayal) the other player was capable of. In other words,

once the anticipated betrayal has been realised and endured, the high BPD trait count individual may be able to engage more cooperatively *knowing* rather than *foreknowing* the risks.

The findings also suggest that individuals high in BPD trait count may be able to repair and/or forgive trust violations, and perhaps even become more trusting after a trust violation. This appears to be at odds with earlier research that found BPD trustees were less likely to engage in coaxing behaviours to win back waning trust (King-Casas et al., 2008). However, this may be a function of power differences in the roles of interest in the studies which were *trustees* in the King-Casas et al. study and *trustors* for the studies comprising this thesis. At the start of each round, it is the trustor who decides whether to send anything to the trustee, therefore holding the power to decide whether to engage the trustee at all. It is also possible that the high BPD trait individuals interpreted the violation differently to their healthier counterparts. During the dissolution phase, the high BPD group most often used negative reciprocity to describe trustee behaviour, while the low BPD group used personal gain. Those with a higher BPD trait count may have viewed the violation by the trustee as a tit-for-tat response to their own ungenerous behaviour during the initial five rounds, and by progressively sending larger amounts during the dissolution phase, the high BPD trait individuals may have been trying to repair the rupture they believed they had instigated. As observed in Study 1, BPD trait count negatively moderated the amount of money invested immediately following the first repair bid by the trustee, which suggests that these individuals may have been coaxing back cooperation during the violation itself, with the first instance of increased reciprocity by the trustee (the first ‘repair bid’) signalling their efforts had been successful and that they could therefore abandon their coaxing efforts.

Lastly, attachment security has been positively associated with trust in various types of relationships and settings (Camgöz & Karapinar, 2016; Fitzpatrick & Lafontaine, 2017; Klest & Philippon, 2016; Mikulincer, 1998). The relationship between attachment insecurity and BPD has well-established theoretical and empirical foundations (Agrawal et al., 2004; Fonagy et al., 1995; Gunderson & Lyons-Ruth, 2008; Linehan, 1993). Based on previous findings it was expected that greater levels of fearful attachment would *amplify* the negative relationship between BPD trait count and trust (Bartz, Simeon, et al., 2011), while greater levels of preoccupied attachment would *diminish* it (Bartz & Lydon, 2006, 2008; Fett et al., 2016). Study 2 revealed as predicted that endorsing a preoccupied attachment style may have tempered the negative effects of BPD on trust formation and repair. However, endorsing a fearful attachment style was associated with a steeper decline in trust during the dissolution phase, which was at odds with the increase in trust associated with BPD trait count. Based on these findings, the paradoxical trust behaviours observed in high BPD trait individuals appear to occur despite the presence of attachment insecurity. However, it is possible that the experimental paradigm was ineffective at activating the attachment system in this study. Thus, it would be premature to totally abandon attachment style as a potential explanatory factor in trust dynamics for those with BPD.

Clinical Implications

Fostering skills to improve interpersonal functioning is a key theme in narratives of personal recovery amongst those with BPD (Ng et al., 2016). Patients describe the capacity to be vulnerable with and trust others as important constituents of relationships (Katsakou et al., 2012). This thesis has contributed to a large body of research on interpersonal dysfunction in BPD (Jeung & Herpertz, 2014; Lazarus et al., 2014; Lis & Bohus, 2013), by uncovering

paradoxical and socially aberrant trust dynamics both in terms of behaviour and reasoning about behaviour, that may contribute to relational dysfunction. While the net cost of adopting the mistrustful and hostile stance in new partner interactions may be one of mutually diminishing cooperation, the overall costs are likely to be much higher including the reinforcement of existing maladaptive belief systems, missed opportunities to experience a sense of social cohesion, and greater difficulty building mutually satisfying relationships. At the most fundamental level, paradoxically punishing cooperative partner behaviour while seemingly rewarding uncooperative partner behaviour may confuse and discourage others.

Clinical interventions that promote reasoning in trust-based interpersonal exchanges is warranted. Mentalization-based therapy aims to develop social-cognitive reasoning about self and others and has been found to improve social and interpersonal functioning (for review see Malda-Castillo et al., 2018), with substantial reductions in interpersonal problems including mistrust observed for patients with BPD (Kvarstein et al., 2015). Moreover, highlighting the difficulties that individuals with BPD have in terms of trust, and acknowledging how trust develops in early childhood, may allow clinicians to offer a non-stigmatizing, non-blaming explanation for relational difficulties. The findings suggest that the topic of trust needs to be addressed explicitly when working with BPD patients. Based on the findings of this thesis and the pre-existing literature, appropriate goals for therapy may include encouraging patients with BPD to reflect on their own capacity to trust and in what ways past experiences may have encouraged a hypervigilant stance; teaching patients to become more aware of their own internal states in interpersonal situations, such as pre-existing feelings-of or expectations-of rejection, and to consider how these might impact their behaviours; to mentalize the thoughts and feelings behind others' trust behaviours; and to

consider how their own, potentially paradoxical behaviours may be perceived by others and consequently responded to.

In addition, trust issues are salient within the therapeutic relationship. Trust has been identified as foundational to the therapeutic relationship among both patients-with and clinicians-treating BPD (Bourke & Grenyer, 2013; Holm & Severinsson, 2011; Langley & Klopper, 2005). Clinicians who work with people with BPD are familiar with the complex and often fraught interpersonal style such individuals exhibit, and the treatment challenges these bring to bear (Bateman & Fonagy, 2004; Clarkin et al., 1999; Kernberg, 1968; Linehan, 1993; Young et al., 2003). For this patient group a strong therapeutic alliance is critical, and clinicians would benefit from being sensitive to the trust difficulties those with BPD face, especially in the early stages of therapy when the patient may be more cautious or uncooperative. Importantly, findings from the studies in this thesis suggest that while borderline pathology may be associated with cautious, mistrustful behaviours at the outset, there appears to be a willingness to repair ruptures, and perhaps even a benefit to having experienced a rupture in the first place. This has important implications for the therapeutic alliance since there is both anecdotal and empirical evidence that rupture and repair experiences may in and of themselves be therapeutic (Eubanks et al., 2018; Safran, 1993; Safran et al., 2011), including for individuals with a personality disorder (Schenk et al., 2020).

Methodological Implications

Despite trust being conceptualised as a dynamic and multiphasic phenomenon (Fulmer, 2010; Fulmer & Gelfand, 2013; Korsgaard et al., 2018; Liebke et al., 2018; Lount et

al., 2008; Schilke et al., 2013; Thayer, 2015), there has been a dearth of research using appropriate methodological and analytical tools to fully exploit its nature (see Korsgaard et al., 2018; Lewicki et al., 2006; Rousseau et al., 1998). When considering the impact of trust on interpersonal functioning and the borderline disposition to rapidly shifting affective states (Nica & Links, 2009), it is limiting to confine measurement of trust to a single point in time, as an average over multiple time points, or to a single phase of trust. Studies 1 and 2 illustrate how in as little as 15 rounds of a TG, trust can be manipulated to examine the effects of variables of interest on how trust forms, dissolves, and restores. The experimental design used offers the benefits of systematizing particular aspects of interpersonal exchanges and offers a valuable adjunct to trust research that uses human dyads (King-Casas et al., 2008; Miano, Dziobek, et al., 2017; Miano, Fertuck, et al., 2017)

In addition to examining whether trust patterns differed between individuals based on their level of BPD symptomatology, the research was also able to examine how trust fluctuated within individuals. That is, intraindividual trust patterns were modelled to look at how BPD trait count and the other covariates modified the rate and magnitude of change within the individual. The notable pattern of intraindividual fluctuations of trust in high trait count individuals was their tendency to take greater interpersonal risks once a person had been proven to be untrustworthy, that is, during the dissolution and restoration phases relative to the formation phase. This paradoxical pattern of trust behaviours set high BPD trait individuals apart from those endorsing few to no BPD traits and supports our earlier assertion that trust violations may perversely result in more trusting behaviours.

The research also made a unique contribution to the BPD trust literature by augmenting the TG which distils trust and cooperation to exchanges in monetary units with

the more qualitative social-cognitive reasoning. This protocol had not been used in BPD trust research, but given the extensive links between BPD and mentalization, understanding how individuals interpret the behaviour of the other player and their own behaviour was a logical next step. The TG paradigm used in the studies comprising this thesis, as well as the data analytical procedures chosen to model trust patterns, has promising potential for future inquiry into trust dynamics in the BPD as well as broader areas of research of interpersonal dynamics in this population.

Limitations

This thesis has made important contributions to the literature as described above, but findings need to be considered in the context of study limitations. Limitations of each of the studies have been included in the discussion sections of the preceding three chapters. The following section provides commentary on the more general limitations of this research. By virtue of experimental design, our research reduced complex and nuanced interpersonal relationships to momentary interactions. As part of our experimental paradigm, we used a simulated trustee programmed to be either unconditionally cooperative or uncooperative, regardless of our participants' behaviour. Although standardizing trustee behaviour enabled us to create uniform conditions for all participants, and to elicit distinct phases of trust, doing so comes at the cost of ecological validity. Interpersonal trust is a dyadic process (Simpson, 2007), and people would be expected to adjust their responses in an interaction based on the behaviour of the person with whom they are interacting. It is questionable whether a human trustee would continue using a positive reciprocity strategy, as was programmed in our TG, in response to the high BPD trait count individual's diminishing investments during the formation phase. The studies in this thesis were inspired by trust research using human dyads (King-Casas et al., 2008; Sharp, Ha, et al., 2011), and it is important to validate and extend

on the current findings by examining whether the trust dynamics observed occur in the same way as in human dyads.

The aim of this body of work was to better understand anomalous trust behaviours in BPD. However, a clinical sample was not used and despite using standardised screening items to identify BPD features, the population drawn from might not reflect a typical patient presentation, limiting the generalizability of the findings. While the severity of interpersonal dysfunction in nonclinical samples of individuals with BPD traits is viewed as approaching that of clinical samples (Tolpin et al., 2004), and persists even when symptoms remit (Trull et al., 1997), some caution should be taken in making inferences about the clinical utility of the findings. Additionally, while the sample size was relatively large, only 7%, or 16 of the 234 participants, met the MSI-BPD's recommended cutoff for clinically relevant levels of BPD symptomatology, a rate nevertheless consistent with population prevalence estimates (APA, 2013). To account for the modest number of high BPD trait count participants, the analysis in Study 1 treated BPD as a dimensional construct, but then repeated and confirmed the findings using cutoffs to create low and high BPD trait count groups. Significantly, even with the low power that accompanies small sample sizes, the effects of interest reached statistical significance. Some of the disadvantages of the smaller sample size were also mitigated by the repeated measures design. Obtaining multiple measures from each participant for each phase of trust maximized power and estimated validity of the analyses, and minimized the estimated error, producing more robust results. Another limitation was the gender imbalance with a predominantly female sample. While the DSM states that BPD is diagnosed in females at a rate of three times that of males (APA, 2013), the prevalence rate as determined by well-controlled studies using representative samples is similar across genders (Torgersen et al., 2001). In response, the effects of gender were controlled for in Study 1 and 2.

The reader should also bear in mind that all three studies were based on a single sample and a single data collection period. At the time of designing the studies that comprise this thesis, we opted to investigate the dynamics of trust behaviours in BPD at a greater depth than had been undertaken in the extant research by introducing multiple innovative methodological and analytical approaches. Generalizability was sacrificed in order to apply these approaches to a single sample providing the advantage of uniformity and allowing us to integrate the findings across all of the studies. In terms of contribution to the literature, we believe that our (a) examination of trust as a dynamic, multiphasic phenomenon; (b) use of DGM to interrogate the changes in trust behaviour; (c) application of qualitative methods to examine the social-cognitive reasoning behind the behaviour; (d) consideration of related variables such as rejection-related affect, self-protective beliefs, and attachment style; and (e) integration of the consequent findings, have together provided a richer understanding of trust dynamics in those with elevated BPD features, and paved the way for more sophisticated and nuanced investigations of not only trust dynamics, but also other interpersonal disturbance that may impact those with BPD. Finally, the TG may capture one type of interdependent encounter, so drawing inferences about other types of relationships based on our findings should be done judiciously. This is particularly relevant when we consider the serious consequences for misplaced trust in other relationships.

Future Directions

A number of recommendations for future directions emerge from some of the limitations described above. In addition, there are several specific areas for future research that should be pursued. The results of this body of research highlight the complex and nuanced relationship between borderline pathology and trust behaviours made prominent in a

simple and anonymous economic exchange game. When we consider the profound interpersonal disturbance that characterizes BPD, and the debilitating effect it has on relationship stability, longevity, and well-being, it is imperative that future research explore trust dynamics in more ecologically valid contexts and in relationships that are more meaningful and have higher stakes. While trust behaviours have been reliably produced under experimental settings with strangers, extrapolating these findings into everyday social and personal interactions requires further investigation using more ecologically valid paradigms. For example, Miano and colleagues used romantic couple dyads to examine how trustworthiness appraisals are modulated by eliciting personal or interpersonal risk (Miano, Dziobek, et al., 2017; Miano, Fertuck, et al., 2017). Combining real life situations in which trust issues may become more salient, such as in online dating, with the means to capture trust-related data, such as daily-diary, experience-sampling, or even an analysis of text-based conversations, may offer more ecologically valid methods with which to explore trust dynamics.

While Studies 2 and 3 increased our understanding of the factors that could underpin how BPD symptomatology may hinder trust behaviours out of the starting gate, neither study was able to provide substantial insight into the paradoxical pattern of trust growth associated with BPD trait count during the dissolution phase. Whether this paradoxical pattern of trust behaviour stems from aberrant social norms, impairments in mentalization, or another yet undetermined cause needs further investigation. Making methodological changes to capture social cognition processes more accurately by eliciting reasoning at the point of sending and receiving monies may provide further clarity around thought processes. This could be done by asking the participant to narrate their decision processes for the investment transactions and their reaction to returns in the moment. Of course, it is possible that eliciting social-

cognitive reasoning in the moment may modify trust behaviours, but this in itself may have meaningful implications for clinical practice.

Future research providing greater clarity around the nature of the self-protective beliefs is also warranted. Self-protective beliefs were measured using three items on the PBQ-BPD (Butler et al., 2002). These items appeared to encompass both a negative belief about others or the relationship and a call to take action to protect oneself. As self-protective beliefs appeared to account for the diminishing trust associated with BPD trait count when initially interacting with a new partner, it is important to augment these findings with more extensive and refined measures of self-protective beliefs and pre-emptive behaviours to understand how these interact together or separately to compromise trust behaviours. For example, it is possible that skills training in how to respond more effectively in interpersonal situations where mistrust or rejection-related concerns are elicited may improve interpersonal outcomes for this population more than seeking to modify trust-related beliefs. Such research will guide the development of interventions designed to improve interpersonal functioning in this population.

Conclusions

The pain and suffering that individuals with BPD experience in the context of interpersonal relationships is profound. While there are invariably a multitude of factors contributing to impaired interpersonal functioning, one of these has been proposed to stem from anomalies in trust including general beliefs about the trustworthiness of others, appraisals of trustworthiness, and trust behaviours. The current thesis adopted a design and analytical methodology to examine trust as a dynamic and multiphasic process and revealed that BPD features were associated with a paradoxical pattern of trust behaviours in which

trusting behaviours declined in the face of cooperative partner behaviour and increased in the face of trust violating partner behaviour. For individuals with a higher number of BPD traits, trust appears to be compromised out of the starting gate, that is, from the beginning of an interaction with a new partner. Such individuals may approach social interactions already feeling rejected, expecting to be betrayed, and be poised to retaliate. They may struggle with understanding the behaviour of others or assume hostile intent, and they may also be unable to explain the reasoning behind their own decisions and those of their partner. The mistrustful behaviours exhibited by high BPD trait individuals are likely to compromise relationships, which may in turn compromise the trajectory of the disorder. Importantly, there appears to be increased trusting behaviour in the face of trust violating actions and trust appears to restore at a faster rate than it was originally formed suggesting ruptures in trust may paradoxically encourage more trusting behaviours in those with BPD features. This phenomenon remains elusive and demands further enquiry, as it has implications for helping individuals with BPD navigate interpersonal difficulties.

Appendices

Appendix A: Published Manuscript: The paradoxical decline and growth of trust as a function of borderline personality disorder: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair (Chapter 2)

Appendix B: Published Manuscript: The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count (Chapter 3)

Appendix C: Research Materials (Chapters 2-4)

Appendix A: Published Manuscript: The paradoxical decline and growth of trust as a function of borderline personality disorder: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair (Chapter 2)

Abramov, G., Miellel, S., Kautz, J., Grenyer, B. F. S., & Deane, F. P. (2020). The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair. *PLOS ONE*, *15*(7).
<https://doi.org/10.1371/journal.pone.0236170>

RESEARCH ARTICLE

The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair

Gamze Abramov^{1*}, Sebastien Mielliet¹, Jason Kautz^{2*}, Brin F. S. Grenyer¹, Frank P. Deane^{1*}

1 School of Psychology, University of Wollongong, Wollongong, NSW, Australia, **2** Darla Moore School of Business, University of South Carolina, Columbia, South Carolina, United States of America

✉ Current address: Naveen Jindal School of Management, University of Texas, Dallas, Texas, United States of America

* ga385@uowmail.edu.au (GA); fdeane@uow.edu.au (FPD)



OPEN ACCESS

Citation: Abramov G, Mielliet S, Kautz J, Grenyer BFS, Deane FP (2020) The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair. *PLoS ONE* 15(7): e0236170. <https://doi.org/10.1371/journal.pone.0236170>

Editor: Christiane Schwieren, Heidelberg University, GERMANY

Received: February 20, 2020

Accepted: June 30, 2020

Published: July 23, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0236170>

Copyright: © 2020 Abramov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Borderline personality disorder (BPD) is associated with paradoxical trust cognitions and behaviours. While BPD is associated with difficulty forming trust and maintaining cooperation in trust-based exchanges, design and analytical methodology best suited to reveal the temporal ebb and flow of trust have been underutilized. We used an economic game to examine the trajectories of trust as it forms, dissolves, and restores in response to trust violation and repair, and to explain how these vary as a function of borderline pathology. Young adults ($N = 234$) played a 15-round trust game in which partner trustworthiness was varied to create three phases: trust formation, trust violation, and trust restoration. Discontinuous growth modelling was employed to capture the trends in trust over time and their relationship with BPD trait count. BPD trait count was associated with an incongruous pattern of trust behaviour in the form of declining trust when interacting with a new and cooperative partner, and paradoxically, increasing trust following multiple instances of trust violation by that partner. BPD trait count was also associated with trust restoring at a faster rate than it was originally formed. By adopting a methodology that recognizes the dynamic nature of trust, this study illustrated at a micro level how relational disturbances may be produced and maintained in those with a moderate to high BPD trait count. Further investigation of the factors and processes that underlie these incongruous trust dynamics is recommended.

Introduction

Borderline Personality Disorder (BPD) is a complex and often enduring disorder with a prevalence rate of 1–2% in the community [1], and 15 to 20% among patients in psychiatric

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This research received financial funding by the University of Wollongong. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

hospitals or outpatient clinics [2]. Disturbed interpersonal functioning has been identified as a core component of BPD in factor analytic studies [3, 4], and is one of the strongest diagnostic discriminators of the disorder [5]. Long-term prospective studies show that even when symptoms remit, improvement in social functioning is limited, with profound and persistent relational impairments [6, 7]. Social network analysis has revealed that individuals with BPD experience a greater number of conflicted relationships and are typically 'cut off' from more people in their networks [8–10], suggesting their relationships are marred by rupture and, potentially, a lack of reconciliation. Since the maintenance of stable partner relationships is associated with recovery [11], it is essential to identify the relational dynamics that contribute to the instability and breakdown of relationships.

Impairments in the capacity to trust has been proposed as a key factor in understanding the relational disturbances of BPD [12]. Believing that others will betray, exploit and deceive is characteristic of the disorder [13, 14]. A systematic review on early maladaptive schemas found the mistrust/abuse schema was one of the most highly endorsed among people with BPD traits or a diagnosis [15]. This suggests that these individuals perceive the world and others as malevolent, hostile, and dangerous. Accordingly, experimental studies using economic exchange games have found that individuals with BPD often behave in ways that compromise the formation and maintenance of trust, particularly once trust has been ruptured [16–18].

Among non-clinical populations, diminished interpersonal trust within intimate relationships is associated with diminished perceptions of relationship quality [19], and that fluctuate markedly within short periods of time [20]. Moreover, breach of trust is associated with relationship dissatisfaction, and is considered to be a relevant factor when deliberating whether to end a relationship [21]. As such, the way that individuals navigate the vicissitudes in trust is likely to impact the quality, stability, and longevity of relationships, and is therefore relevant to improving interpersonal functioning in BPD. Using an economic game paradigm, the current study seeks to examine how the presence of BPD traits modifies interpersonal trust processes, including how trust changes in response to violation and repair.

The Trust Game (TG)

The experimental paradigm most frequently used to examine trust behaviours is the trust game [TG: 22]. Typically, in a two-player game, one person—the *investor*—is allocated an endowment (e.g. \$10) and can choose whether to entrust any of it with the other player—the *trustee*—for investment. The amount invested is automatically multiplied by a factor—most often three—before being received by the trustee. The trustee can then reciprocate if they so desire, by sending the investor a sum of their choosing from this tripled amount. Trust is operationalized as the proportion of the original endowment transferred by the investor. The TG can be played with a human dyad or a programmed agent, the latter allowing researchers to manipulate responses such as the magnitude of the sum repaid, to investigate how trust changes over time and in response to discrete events.

Interpersonal trust as a dynamic phenomenon

While the TG is an ideal vehicle for examining trust dynamically, the methodological and data analytical procedures best suited to reveal the dyadic ebb and flow of trust and cooperation across repeated interactions have been underutilized [see 23–25], with a few exceptions within the organizational psychology sphere [e.g., 24, 26].

Trust as a temporal phenomenon has been conceptualized as comprising at least three distinct phases: formation, dissolution, and restoration [18, 24, 27–31]. Trust formation refers to the development of trust in a new relationship; trust dissolution refers to the decline of trust in

response to a violation of one's trust by another party; and trust restoration refers to the rebuilding of trust subsequent to trust dissolution, in response to reparative attempts by the offending party. Given that individuals with BPD behave in ways that jeopardise trust development and maintenance [16–18], it is important to conduct research that captures its temporal and mutable nature.

Trust and BPD

In the BPD literature, trust has not typically been treated as a dynamic and multiphasic construct. Rather, it has been measured at a single point in time [32], aggregated across multiple time points [18, 33], or restricted to focus on a single phase of trust [17]. To the authors' knowledge, only one study has accounted for trust dissolution and trust restoration by manipulating trustee reciprocity to differentiate between three phases: cooperative reciprocity, where the trustee returned a profit across five consecutive rounds; trust rupture, where the trustee kept the entire investment; and trust repair, where the trustee behaved cooperatively following rupture [18]. Trust was operationalized as the amount transferred by the investor to the trustee for investment, averaged across the aggregated rounds of each phase. Liebke, Koppe [18] found no difference between BPD patients and healthy controls (HC) in the average amount invested during the rupture or repair phases but found evidence that patients may transfer less during the cooperative phases, what might loosely be defined as the trust formation phase.

The latter findings were consistent with a 5-round TG study in which feedback on trustee reciprocity was withheld [17]. Results indicated that average levels of trust were lower for patients with BPD than patients with major depressive disorder (MDD) or HC's. Linear trend analyses further revealed that while both control groups progressively increased their investments over the five rounds, growth was not observed for the BPD group. Together these findings suggest that individuals with BPD have a greater propensity towards mistrust when interacting with a new partner, even in the absence of investment loss.

The temporal nature of trust in BPD was also examined in a study focusing on trustee rather than investor behaviour. King-Casas, Sharp [16] used human dyads to elegantly quantify the process of trustee cooperation in negotiating trust rupture and repair by pairing HC investors with either HC or BPD trustees in a 10-round TG. They discovered that in response to declining investments by the investor—a signal of diminishing trust—healthy trustees increased their reciprocity over subsequent rounds, effectively 'coaxing' the investor to send larger investments. In contrast, BPD trustees were more likely to respond by reducing reciprocity further or keeping the entire investment, resulting in increased incidences of breakdown in cooperation with investors sending even less or nothing in subsequent rounds. They also found differential neurological activity in the bilateral anterior insula between BPD and HC trustees following receipt of a smaller investment. This area is associated with defection in social exchanges following norm violations [for reviews see 34, 35], leading King-Casas and colleagues' to propose that atypical social norms may underlie the reduced incidence of coaxing behaviours among BPD trustees, by way of a failure to recognize and/or respond appropriately to the social signals of reduced trust.

Although King-Casas and colleagues' study measured trustee cooperation rather than trust, their findings compliment the BPD trust literature to suggest that individuals with BPD compromise the maintenance of trust by not acting to preserve trust when a rupture has occurred [16]. Considered together, the extant research highlights the anomalous relationship between BPD and interpersonal trust, but also exposes gaps in the research, particularly in understanding how trust dissolves and is rebuilt in relation to BPD. Considering this, the current study

has paired the TG with appropriate methodological and data analytical procedures better suited to address the dynamic and multiphasic nature of trust.

Measuring change with Discontinuous Growth Modelling (DGM)

In recent years, researchers in the organizational psychology field have advanced the research into trust as a dynamic process by employing longitudinal designs and data analytic approaches suitable to capture how trust unfolds across the various phases of trust [24, 26, 27]. These studies have utilized discontinuous growth modelling [DGM: 36, 37], a derivative of mixed effects modelling that can model longitudinal data whilst accounting for discontinuities in the data such as an experimenter-induced trust violation event. This methodological and analytical pairing has enabled examination into the individual differences and higher-level factors that influence how trust forms, dissolves, and restores.

In addition to allowing researchers to measure changes in trust over time and in response to specific events, the change parameters can be coded to measure either *absolute* change in trust or the *relative* intraindividual fluctuations in trust. While using relative coding is the standard in DGM [e.g., 37, 38], Bliese and Lang recommend that modelling both types of change may better inform theory development and practical applications [36]. For example, how quickly trust is rebuilt in *absolute* terms can be determined by comparing the trajectory of trust during the trust restoration phase to zero, and further test whether this trajectory is modified by the presence of BPD symptoms. On the other hand, how quickly trust is rebuilt *relative* to how quickly it was initially built can be determined by comparing the trajectory of trust during restoration to the trajectory of trust before the violation took place. As a second step, it is then possible to determine whether BPD moderates the *relative* difference in the trajectories. Examining relative change is especially important if there is a notable linear trend during the period before a discontinuity. The comparatively slower rate of trust growth observed in BPD patients in the initial stages of an exchange [17] supports the inclusion of relative coding.

Current study and aims

The current study uses a multi-round TG and discontinuous growth modelling to examine how BPD impacts interpersonal trust processes, including the development of trust with a new partner, how trust dissolves in response to trust violation, and the restoration of trust in response to trust repair. Based on the methodology adopted in previous works [24, 26, 27], a 15-round TG is used with modified reciprocity to precipitate a distinct trust violation, before resuming a pre-violation rate of reciprocity to precipitate a trust repair. These variations in reciprocity are intended to elicit the three trust phases of formation, dissolution, and restoration. This allows five change parameters to be modelled: three parameters representing the rate of trust growth within each of the phases and two parameters representing the change in the level of trust between the phases.

The current study primarily aims to examine whether the number of BPD traits reported modifies, (a) how much trust decreases immediately following the violation (dissolution transition), and how much trust increases immediately following the repair (restoration transition), and (b) the direction and rate in which trust changes in each phase (formation slope, dissolution slope, and restoration slope). Based on previous findings that BPD is associated with increased mistrust in the early stages of a social exchange [17, 18] and a reduced tendency to utilize trust reparative behaviours to maintain cooperation in a social exchange [16], it is hypothesized that BPD trait count will be associated with a more pronounced decrease in trust after the initial instance of violation, a less pronounced increase in trust after the initial

instance of repair, and a deleterious effect on the trajectories of trust within each of the formation, dissolution, and restoration phases.

Our secondary aims are to examine the fluctuations in trust at an intraindividual level and to determine whether BPD trait count modifies the overall pattern of these fluctuations. In order to determine how the number of BPD traits moderates the rate of trust growth during the dissolution and restoration phases, we switch our focus from *absolute* differences as described in the primary aim, to *relative* differences. Specifically, the study will examine (a) the rate that trust dissolves *relative* to how quickly it was originally built (i.e. dissolution phase vs formation phase); (b) the rate trust is rebuilt *relative* to how quickly it dissolved (i.e. restoration phase vs dissolution phase); and (c) the rate trust is rebuilt *relative* to how quickly it was originally built (i.e. restoration phase vs formation phase).

The uncooperative behaviour observed among the trustees with BPD in King-Casas, Sharp [16] did not appear until the latter half of the TG, suggesting their ability to maintain a trust relationship deteriorated over time. Given that BPD traits are expected to be associated with a slower rate of trust growth during the formation phase, it is hypothesized that even when this initial pattern of trust change is taken into account, trust growth during the latter trust phases will be even slower for those with a high number of BPD traits. In other words, as BPD trait count increases, the trajectory of monetary units transferred during dissolution and/or restoration in comparison to the trajectory of transfers made during formation, will be more negative.

The final hypothesis refers to changes in the rate and direction of change during restoration relative to dissolution. Based on King-Casas et al. [16] finding that trustees with BPD were less likely to coax by increasing reciprocity in order to maintain cooperation, it is possible that in a similar vein, investors with high levels of BPD traits may not respond as favourably to the increase in reciprocity levels in regards to increasing their rate of trust growth. It is hypothesized that even after taking into account the rate of trust dissolving in the dissolution phase, BPD trait count will be associated with a slower rate of trust growth during restoration.

Cognitive reflection—the act of problem solving by engaging in conscious deliberation and suppressing intuitive/impulsive responding [39]—has been positively associated with trust in TG's [40]. Additionally, as BPD is associated with impairments in executive functioning [see 41, 42], altered decision making [see 43], deficits in social problem-solving [see 44], and impulsivity in interpersonal contexts [45], including a measure to control for cognitive reflection may be warranted. In regards to gender, a meta-analysis found that female investors invest significantly less than males in the TG [46]. As BPD is diagnosed at a rate of 3:1 in females compared to males [APA, 47], accounting for the effects of gender when examining the influence of BPD traits on trust is justified. Consequently, in both the absolute and relative models, the main effects of cognitive reflective ability and gender will be controlled.

Materials and methods

The study was approved by the University of Wollongong ethics committee HE2017/253. All participants provided written informed consent.

Participant

Participants were undergraduate students from a large Australian university who elected to take part in a psychology research participation program in exchange for course credit. As part of an additional study, after playing the TG, participants were asked to describe their own and their partner's intentions in reference to specific transactions made. Participants whose responses indicated scepticism that their partner was human (and not a computer algorithm)

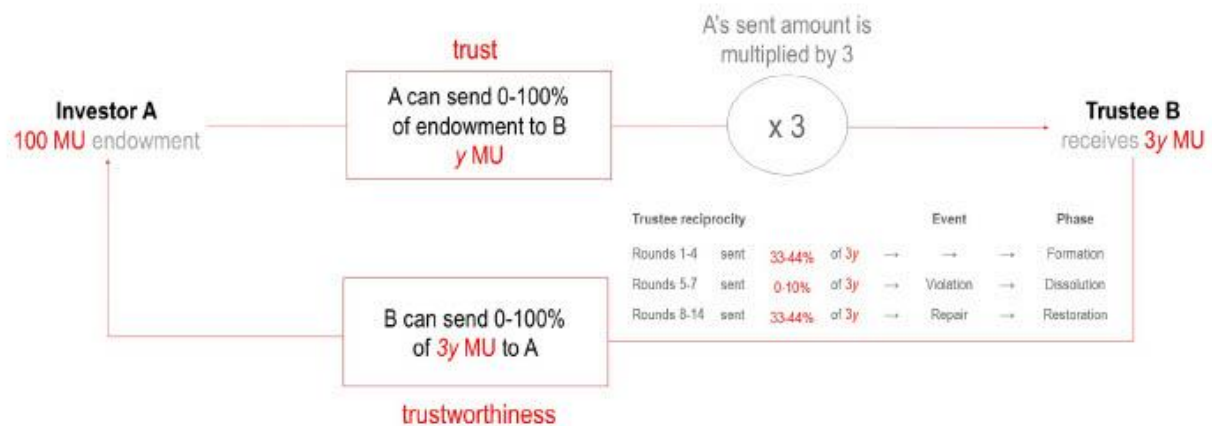


Fig 1. Trust game structure of a single round. At the start of each round, investors were allocated 100 monetary units (MUs) from which they could send any amount to the trustee for investment. After receiving the transferred amount multiplied by a factor of 3 (3yMU), the trustee could then return any amount to the investor. Trustee reciprocity was randomized to fall between 33–44% and 0–10% to induce violation and repair, respectively.

<https://doi.org/10.1371/journal.pone.0236170.g001>

were dropped from the final analyses ($n = 7$), leaving a final sample of 234 (64% female; $M_{age} = 20.87$, $SD_{age} = 5.66$). The recruitment information advised participants that the online study was looking at the relationship between economic decision making and various personality variables and would involve questionnaires and playing an investment game.

Trust Game (TG) protocol

The current study used a multi-round TG comprising of 15 sequential rounds played with the same partner. At the beginning of each round the investor was allocated 100 monetary units (MU) by the experimenter and given the option to send the trustee any proportion from 0–100% for investment. The amount sent was automatically tripled, and the trustee could repay any proportion from 0–100% of the tripled investment to the investor. During rounds 1–4 and 8–14 (inclusive), trustee repayments were randomized to fall between 34–44% of the tripled investment, providing the investor with a return the equivalent of the original investment plus up to 32% profit (range 0–32%). A trustee reciprocity range of 34–44% was selected to ensure repayments would be comparable to the reciprocity ratios observed in human trustees in previous research [16, 46]. During rounds 5–7 (inclusive), trustee repayments were randomized to fall between 0–10% of the tripled investment, providing the investor with a return the equivalent of losing from 70% to 100% of their original investment (due to rounding to the nearest whole number, investments of 1 MU did not incur a loss). This loss was designed to induce a trust violation, and it was repeated over three consecutive rounds to reinforce the participant's perception that the violation was deliberate and not construed as a mistake [27, 48]. Based on this repayment schedule, investments can be divided into three distinct trust phases: formation (rounds 1–5), dissolution (rounds 6–8), and restoration (rounds 9–15). Fig 1 illustrates the process of a single exchange and the rates of trustee reciprocity preceding each event. Following each round, the participant was provided with a summary indicating how much they invested, how much was repaid, and the final sum each party ended up with for that round. As participants saw the summary at the end of each round, trust rates lag trustee reciprocity rates by 1 round.

Procedure

Participants registered and completed the study online during the study period. After completing a series of questionnaires, participants were given instructions on how to play the TG and were required to correctly answer three test questions to demonstrate their understanding of the game before being permitted to proceed. Despite being told roles would be allocated randomly, participants were all assigned the role of investor. They were also informed that they would be paired with another study participant from a participating academic institution, when in fact the other 'player' was a computer programmed trustee. This form of deception has been used in most of the experiments using the same protocol in a BPD population [e.g., 17, 33]. The use of a computerized agent rather than a human confederate allows standardization of trustee responses in terms of reciprocity levels and response time. Participants were not advised how many rounds they would play as previous research has shown that the defection rate increases when people know when a social exchange will end [49]. Participants were not offered a monetary incentive to participate but at the conclusion of the study, they were automatically allocated course credit.

Measures

Trust. The number of MU that participants entrusted to their partner in each round, ranging from 0 to 100, represents a single behavioural measure of trust. Each participant provided 15 trust measures in total.

BPD trait count. The McLean Screening Instrument for Borderline Personality Disorder [MSI-BPD; 50] is a 10-item self-report screening instrument for BPD, with demonstrated internal consistency, validity, test-retest reliability, and in individuals aged 18 to 25, excellent sensitivity (.90) and specificity (.93) for the diagnosis of DSM-IV BPD [APA, 51]. The MSI-BPD has satisfactory internal consistency (Cronbach's $\alpha = .74$, $N = 200$) and good test and retest reliability (Spearman's $\rho = .72$) [50]. In the current study the MSI-BPD was used as a measure of BPD trait count and provided a score ranging from 0 to 10.

In our sample the MSI-BPD demonstrated very good internal consistency (Cronbach's $\alpha = .84$, $N = 234$). The MSI-BPD additionally had significant moderate positive correlations with the Mental Health Inventory [MHI-5; 52], a measure of general mental health and quality of life ($r = .54$, $p < .001$), and the Standardized Assessment of Personality—Abbreviated Scale Self-Report [SAPAS-SR; 53], a measure of general personality psychopathology ($r = .54$, $p < .001$). In the current sample, 24% reported 3 or more BPD traits, a level of symptomatology that is considered to reflect the presence of a notable level of borderline pathology [CLPS: 6]. Finally, using the MSI-BPD conservative cut-off of 7/10, 16 participants met caseness for BPD.

Cognitive reflective ability. The Cognitive Reflection Test [CRT; 39] is a three-item measure of the willingness to engage in deliberation during a cognitive task. Each item is a deceptively simple mathematical problem in which there is an intuitive but incorrect answer. The CRT aims to measure the ability or disposition to resist responding impulsively by assessing the number of questions answered correctly. Participants are required to answer each question by typing the correct response in digits. Correct responses receive a score of 1 with all other responses scored 0, resulting in a total summed CRT score ranging from 0–3. The CRT has a moderate positive correlation with measures of intelligence and is correlated highly with various measures of decision-making indices [39]. In the current study 48% of participants scored 0, 18% scored 1, 17% scored 2, and 17% scored 3, with an overall sample mean of 1.03, which falls within the range of CRT scores collected from other academic institutions [see 39].

Post-game trustee appraisals. After the game participants rated the trustee on fairness (*‘Did the other player play fair?’*) and trustworthiness (*‘Is the other player trustworthy?’*) on 5-point Likert-type scales (0 = “Not at all”; 4 = “Absolutely”).

Data analyses. The NLME package [54], included in the open source software R [55], was used to conduct discontinuous growth modelling (DGM) analyses [see 36, 37]. We tested two-level mixed-effects models, with investment occasions (rounds) at Level 1 nested within individuals at Level 2. Level 1 parameters were coded based on the framework recommended by Singer and Willett [37] and Bliese and Lang [36] to examine changes in the average level of trust between and growth within each of the formation (pre-violation), dissolution, and restoration phases. Five change variables were created to examine *absolute* and *relative* change. According to Bliese and Lang [36], the coding and combination of these change variables into a matrix allows for the regression coefficients to describe the change in the dependent variable in different ways. As we were interested in how individuals respond first to a trust violation and then to a trust restoration, we coded the change variables such that the coefficients reflect the previous stage as the baseline for interpretation. Specifically, the change variables coded for the dissolution phase (DT and DS) use the formation phase as a baseline and change variables coded for the restoration phase (RT and RS) use the dissolution phase as a baseline. The coding and interpretation for each change variable are presented in [Table 1a and 1b](#) for absolute and relative change, respectively. This coding allowed for easy interpretation of how individuals responded during the three phases of formation, dissolution, and restoration. Finally, an alternative coding system was used to reflect relative change using the formation phase as a baseline (see [Table 1c](#)).

We initially examined Level 1 change by including only Level 1 predictors in a series of models to calculate a basic DGM. Next, we examined the Level 2 model by including BPD trait count as a Level 2 predictor to account for differences in Level 1 change, while also controlling for the main effects of gender and cognitive ability. Snijders and Bosker [56] maintain that as a result of reduced parameter reliability in multilevel analysis, the power to detect cross-level interactions may be insufficient (p. 94). To account for this, a criterion level of $p < .10$ was used for all cross-level interactions effects, and $p < .05$ for all other effects [see also 57, 58], and all tests conducted were two-tailed. We tested a model examining *absolute* change to address our first research aim followed by models examining *relative* change to address our second research aim. For the mixed-effects analyses, all level 2 predictors were *z*-standardized and centered at the sample mean.

Results

Descriptive data and intercorrelations

The dataset for the current study can be accessed in [S1 Dataset](#). [Table 2](#) presents the means, standard deviations, and intercorrelations for BPD trait count, MUs transferred within each trust phase, partner appraisals, and cognitive reflection.

BPD trait count had a significant negative correlation with MUs transferred during formation but was not found to be significantly associated with MUs transferred during the dissolution or restoration phases. BPD trait count was not significantly associated with appraisals of trustworthiness or fairness. However, trustworthiness appraisals were positively associated with the number of MUs transferred during all three trust phases while fairness appraisals were positively associated with the amount transferred during the dissolution phase only.

To examine whether there were systematic differences in trustee reciprocity rates we conducted both correlation and ANOVA analyses so that we could treat BPD trait count as a continuous and categorical variable. We created three BPD categories based on number of traits

Table 1. Coding and interpretation of change variables in the discontinuous mixed-effects growth models^a.

a) Absolute change																
Change variable	Formation														Dissolution	
	Restoration														Interpretation of associated parameter estimates in the DGM	
Rounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
TIME _H	0	1	2	3	4	4	4	4	4	4	4	4	4	4	4	Linear change of MUs transferred in the formation phase (π_{1i})
DT _H	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	Difference in level of MUs transferred immediately following the trust violation (π_{2i}), Round 6 vs Round 5
DS _H	0	0	0	0	0	0	1	2	2	2	2	2	2	2	2	Linear change of MUs transferred in the dissolution phase
RT _H	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	Difference in level of MUs transferred immediately following the trust repair (π_{4i}), Round 9 vs Round 8
RS _H	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	Linear change of MUs transferred in the restoration phase
b) Relative change (relative to preceding phase)																
Change variable	Formation														Dissolution	
	Restoration														Interpretation of associated parameter estimates in the DGM	
Rounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
TIME _H	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Linear change of MUs transferred in the formation phase (π_{1i}), also referred to as the pre-violation period
DT _H	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	Difference in MUs transferred immediately following the trust violation (π_{2i}) relative to the predicted transfer amount based on the formation phase (i.e. the expected MUs at Round 6 in the absence of trust violation)
DS _H	0	0	0	0	0	0	1	2	3	4	5	6	7	8	9	Linear change of MUs transferred in the dissolution phase relative to the formation phase (i.e. the pre-violation period) (π_{3i})
RT _H	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	Difference in MUs transferred immediately following the trust repair (π_{4i}) relative to the predicted transfer amount based on the dissolution phase (i.e. the expected MUs at Round 9 in the absence of trust repair)
RS _H	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	Linear change of MUs transferred in the restoration phase relative to the dissolution phase (i.e. the pre-repair period) (π_{5i})
c) Relative change (relative to formation phase)																
Change variable	Formation														Dissolution	
	Restoration														Interpretation of associated parameter estimates in the DGM	
Rounds	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
TIME _H	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Linear change in MUs transferred in the formation phase (π_{1i}), also referred to as the pre-violation period
DT _H	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	Difference in MUs transferred immediately following the trust violation (π_{2i}) relative to the predicted transfer amount based on the formation phase (i.e. the expected MUs at Round 6 in the absence of trust violation)
DS _H	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	Linear change in MUs transferred in the dissolution phase relative to the formation phase (i.e. the pre-violation period) (π_{3i})
RT _H	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	Difference in MUs transferred immediately following the trust repair (π_{4i}) relative to the predicted transfer amount based on the formation slope (i.e. the expected MUs at Round 9 in the absence of trust violation/repair)
RS _H	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	Linear change in MUs transferred in the restoration phase relative to the formation phase (i.e. the pre-violation period) (π_{5i})

Rounds = measurement occasions in the trust game, TIME_H = linear change, DT_H = dissolution transition, DS_H = dissolution slope, RT_H = restoration transition, RS_H = restoration slope

^a As this is a complex coding scheme, we asked several experts in the use of DGM to evaluate and ensure the interpretation of the variables based on the matrices presented. See Acknowledgments.

<https://doi.org/10.1371/journal.pone.0236170.t001>

Table 2. Intercorrelations, means, and standard deviations of study variables.

	1	2	3	4	5	6	7
1. BPD	-						
2. Formation	-0.18**	-					
3. Dissolution	-0.01	0.34***	-				
4. Restoration	-0.08	0.44***	0.55***	-			
5. Trustworthiness	-0.10	0.19**	0.16*	0.15*	-		
6. Fairness	-0.10	0.04	0.15*	0.07	0.49***	-	
7. Cognitive reflective test	-0.11†	0.23***	0.03	0.13†	0.03	0.03	-
Mean	1.63	46.16	24.54	38.4	2.12	2.71	1.03
SD	2.34	18.95	19.14	23.29	1.08	1.18	1.15

† $p < .10$,* $p < .05$,** $p < .01$,*** $p < .001$ $n = 234$ participants. Spearman correlation.

BPD traits (0–10). Formation/Dissolution/Restoration (0–100). Trustworthiness/Fairness appraisal (1–5). Cognitive reflective test (0–3).

<https://doi.org/10.1371/journal.pone.0236170.t002>

endorsed on the MSI-BPD: Low (0–2), Moderate (3–6), and High (7–10). Due to significant discrepancies in sample size ($N = 177/41/16$, respectively) and evidence of non-normality and heterogeneity of variance in some distributions, we elected to run a non-parametric ANOVA analysis. Both, the correlation approach and the Kruskal–Wallis H Test did not show any significant difference in trustee reciprocity between BPD categories.

Cognitive functioning had a marginally significant negative correlation with BPD trait count and was positively associated with MUs transferred during formation and dissolution (see Table 1). T-tests suggest that transfers made during formation differ by gender ($p < .01$), with females ($M = 43.75$, $SD = 17.77$) transferring fewer MUs than males ($M = 50.38$, $SD = 20.27$), and during restoration ($p < .001$), with females ($M = 34.39$, $SD = 22.12$) transferring fewer MUs than males ($M = 45.43$, $SD = 23.75$). Females ($M = 23.56$, $SD = 18.42$) and males ($M = 26.24$, $SD = 20.35$) transferred comparable MUs during dissolution ($p = ns$). T-tests also indicated a marginally significant difference in BPD trait count by gender ($p = .06$), with females ($M = 1.85$, $SD = 2.46$) reporting a slightly higher number of traits than males ($M = 1.25$, $SD = 2.08$).

Individual differences in trust patterns

We began by testing the random intercept model (null model) to estimate the intraclass correlation coefficient (ICC) to determine how much of the variability in MUs transferred across the 15 rounds resulted from between-person differences. The ICC was .253, indicating 25.3% of the variance in the amount invested across rounds can be explained by properties of the individual. This ICC value is consistent with our expectations based on prior exploration of trust behaviour during the trust game [24] and the knowledge that our experimental design possessed three distinct trust phases.

Level 1 analyses

Following the procedure established by Bliese and colleagues [36, 59], for each of our absolute and relative analyses, we first generated a linear-only baseline DGM (random intercept model)

Table 3. Model comparison tests for discontinuous growth models, auto correlations, and heteroscedasticity.

Model	df	AIC	BIC	logLik	Test	L.Ratio
a) Absolute model						
1. Random Intercept Model	8	33229.79	33279.08	-16606.89		
2. Random TIME	10	33206.92	33268.54	-16593.46	1 vs 2	26.87***
3. Random TIME & DT	13	33114.75	33194.85	-16544.37	2 vs 3	98.17***
4. Random TIME, DT & DS	17	33072.36	33177.11	-16519.18	3 vs 4	50.39***
5. Random TIME, DT, DS & RT	22	33057.37	33192.92	-16506.68	4 vs 5	24.99***
6. Random TIME, DT, DS, RT & RS	28	32991.98	33164.50	-16467.99	5 vs 6	77.39***
7. Autocorrelation Error Structure	29	32968.31	33147.00	-16455.15	6 vs 7	25.67***
8. Heteroscedasticity	29	32992.60	33171.29	-16467.30	6 vs 8	1.38
b) Relative model (preceding phase as a baseline)						
1. Random Intercept Model	8	33229.79	33279.08	-16606.89		
2. Random TIME	10	33113.18	33174.80	-16546.59	1 vs 2	120.60***
3. Random TIME & DT	13	33065.67	33145.77	-16519.83	2 vs 3	53.51***
4. Random TIME, DT & DS	17	33033.18	33137.93	-16499.59	3 vs 4	40.49***
5. Random TIME, DT, DS & RT	22	33002.31	33137.87	-16479.16	4 vs 5	40.87***
6. Random TIME, DT, DS, RT & RS	28	32992.08	33164.61	-16468.04	5 vs 6	22.23***
7. Autocorrelation Error Structure	29	32968.45	33147.14	-16455.22	6 vs 7	25.63***
8. Heteroscedasticity	29	32992.61	33171.30	-16467.31	6 vs 8	1.47
c) Relative model (formation phase as a baseline)						
1. Random Intercept Model	8	33229.79	33279.08	-16606.89		
2. Random TIME	10	33113.18	33174.80	-16546.59	1 vs 2	120.60***
3. Random TIME & DT	13	33103.59	33183.70	-16538.80	2 vs 3	15.59**
4. Random TIME, DT & DS	17	33109.18	33213.93	-16537.59	3 vs 4	2.42
5. Random TIME, DT, DS & RT	22	33037.07	33172.63	-16496.54	4 vs 5	82.11***
6. Random TIME, DT, DS, RT & RS	28	32992.06	33164.58	-16468.03	5 vs 6	57.01***
7. Autocorrelation Error Structure	29	32968.41	33147.10	-16455.21	6 vs 7	25.65***
8. Heteroscedasticity	29	32992.67	33171.35	-16467.33	6 vs 8	1.39

**p* < .05,

***p* < .01,

****p* < .001.

DT = dissolution transition, DS = dissolution slope, RT = restoration transition, RS = restoration slope.

<https://doi.org/10.1371/journal.pone.0236170.t003>

to determine the pattern of change in trust for participants as a whole using the TIME variable.

$$Trust_{it} = \pi_{0i} + \pi_{1i}TIME_{it} + \pi_{2i}DT_{it} + \pi_{3i}DS_{it} + \pi_{4i}RT_{it} + \pi_{5i}RS_{it} + \epsilon_{it}$$

This model consisted of an intercept (π_{0i}), error variance (ϵ_{it}), and the change variables: linear change over time ($TIME_{it}$), dissolution transition (DT_{it}), dissolution slope (DS_{it}), restoration transition (RT_{it}), and restoration slope (RS_{it}), as described in Table 1.

We then compared the random intercept model, which describes the overall trajectory averaged across all participants, with a model that allows the trajectory of each participant to vary. By allowing for random variance across each change coefficient, we can estimate whether there are between-person differences in the pattern of MUs transferred. Beginning with the random intercept model, each consecutive model allows an additional change coefficient to freely vary across participants. Each consecutive model was tested against the previous model using the log-likelihood test (for each model, see rows 1 to 6 of Table 3a–3c below). Results indicate that a model accounting for random effects for all change coefficients was the best

fitting model for both the absolute and relative models, except for the dissolution slope in the model looking at change relative to the formation phase (see Table 3c). However, this does not affect the interpretation of this model since its purpose in the present study is to examine trust changes during the restoration phase in comparison to trust levels and growth before the violation took place.

In the next step, we tested for lag-1 autocorrelation (row 7 of Table 3 for each model) and heteroscedasticity (row 8 of Table 3 for each model). Log-likelihood ratio tests indicated a significantly better fit only when we accounting for autocorrelation. Models that controlled for both error structures simultaneously did not converge. The Final DGM's in Table 4 below provides the parameter estimates and standard errors with random transitions and slopes, and autocorrelation included.

The results of the absolute model indicate that, on average, the MUs transferred decreased significantly immediately following the trust violation ($est_{DT} = -20.60, p < .001$) and increased significantly immediately following the trust repair ($est_{RT} = 10.83, p < .001$). This confirms that the experimental manipulation of trustee reciprocity to create trust violation and trust restoration was successful. There was no evidence of a linear trend in MUs transferred during the formation ($est_{PS} = -.12, n.s.$) and dissolution phases ($est_{DS} = -.80, n.s.$), but there was a positive linear trend in MUs transferred during the restoration phase ($est_{RS} = 1.26, p < .01$). The results of the corresponding relative model further indicate that the immediate shift in MUs transferred following the trust violation ($est_{DT} = -20.49, p < .001$) resulted in transfer amounts that were significantly lower than what would be expected if the violation had not occurred. Furthermore, the immediate shift in the MUs transferred following the trust repair ($est_{RT} = 11.63, p < .001$) resulted in transfer amounts that were significantly higher than what would be expected if the repair had not occurred. However, when using the formation phase as a baseline, this increase in transfer amounts at repair was still significantly lower than what would be expected if neither the violation nor repair had occurred ($est_{RT} = -10.91, p < .01$). Regarding trust growth, the slope for the dissolution phase was not found to differ significantly from the slope during the formation phase ($est_{DS} = -.69, n.s.$). However, the positive linear trend in MUs transferred during restoration was steeper than that observed during the formation phase ($est_{RS} = 1.38, p < .10$) and during the dissolution phase ($est_{RS} = 2.07, p < .10$).

Level 2 analyses

To test for systematic differences in MUs transferred between individuals due to differences in the number of BPD traits reported, BPD trait count was added as a Level 2 predictor for each

Table 4. Final discontinuous growth models.

Variables	Absolute		Relative (to preceding phase)		Relative (to formation phase)	
	Est	SE	Est	SE	Est	SE
Intercept	46.42**	1.51	46.42***	1.51	46.42***	1.51
Formation slope (FS)	-0.12	0.58	-0.12	0.58	-0.12	0.58
Dissolution transition (DT)	-20.60**	2.39	-20.49***	2.79	-20.49***	2.79
Dissolution slope (DS)	-0.80	1.16	-0.69	1.28	-0.69	1.28
Restoration transition (RT)	10.83**	1.87	11.63**	2.63	-10.91**	4.15
Restoration slope (RS)	1.26**	0.40	2.07†	1.25	1.38†	0.71

† $p < .10$,

* $p < .05$,

** $p < .01$,

*** $p < .001$, tests are two-tailed, $n = 234$ participants, 3510 observations.

<https://doi.org/10.1371/journal.pone.0236170.t004>

Table 5. Discontinuous mixed-effects growth models predicting trust as a function of BPD trait count after controlling for gender and cognitive reflective ability.

	Absolute		Relative (preceding phase)		Relative (formation phase)	
	Est	SE	Est	SE	Est	SE
Intercept	49.14**	2.02	49.12***	2.02	49.20***	2.02
Formation slope (FS)	-0.08	0.57	-0.08	0.57	-0.08	0.57
Dissolution transition (DT)	-20.74**	2.40	-20.67***	2.79	-20.67***	2.80
Dissolution slope (DS)	-0.76	1.16	-0.69*	1.27	-0.69	1.27
Restoration transition (RT)	10.55**	1.86	11.31**	2.63	-11.42**	4.12
Restoration slope (RS)	1.29**	0.40	2.05	1.26	1.36	0.71
Gender (Female)	-4.30*	2.15	-4.28*	2.15	-4.41*	2.15
Cognitive reflective ability	2.34*	1.04	2.34*	1.04	2.36*	1.03
BPD	0.17	1.49	0.17	1.49	0.18	1.50
FS * BPD	-1.12 [†]	0.57	-1.12 [†]	0.57	-1.12 [†]	0.57
DT * BPD	2.13	2.40	3.25	2.79	3.25	2.79
DS * BPD	2.33*	1.16	3.45**	1.27	3.45*	1.27
RT * BPD	-3.83*	1.85	-6.16*	2.62	7.43 [†]	4.11
RS * BPD	0.43	0.40	-1.90	1.25	1.55*	0.71

[†]*p* < .10,

**p* < .05,

***p* < .01,

****p* < .001, tests are two-tailed, *n* = 234 participants, 3510 observations. BPD = Number of BPD traits reported on MSI-BPD. BPD and cognitive reflective ability were z-standardized and centered at the sample mean.

<https://doi.org/10.1371/journal.pone.0236170.t005>

of the Level 1 components. The associated Level 2 equations are as follows:

$$\begin{aligned} \pi_{0i} &= \beta_{00} + \beta_{01}(BPD)_i + r_{0i} \\ &\vdots \\ \pi_{5i} &= \beta_{50} + \beta_{51}(BPD)_i + r_{5i} \end{aligned}$$

Based on results from separate main effects models for gender (*est*_{Female} = -5.64, *p* < .01) and cognitive reflective ability (*est*_{Cognitive} = 2.52, *p* < .01) which found that being female or having low cognitive reflective ability were associated with smaller transfers overall, we controlled for both in our final interaction models. Results for the final discontinuous mixed-effects models are presented in Table 5.

Fig 2 graphs the effects of BPD on the overall change pattern of MUs transferred. MU transferred at each measurement occasion was predicted for individuals with a high (1 SD above the sample mean) and low (1 SD below the sample mean) BPD trait count, contrasted with predicted MUs transferred for individuals scoring at the sample mean of BPD. Results for the absolute model are reported first, followed by an outline of any notable deviations observed when looking at relative change.

Results do not indicate a main effect of BPD trait count. With respect to changes in MUs transferred, it was hypothesized that BPD trait count would have a deleterious effect on the amount transferred at each transition, representing acute changes in trust in reaction to violation and repair, and on the rate of change of MUs transferred during each phase, representing the rate of trust growth or decline.

Formation slope. We predicted that BPD would be associated with a slower rate of trust growth during the formation phase. As indicated by Table 5, this hypothesis was confirmed for

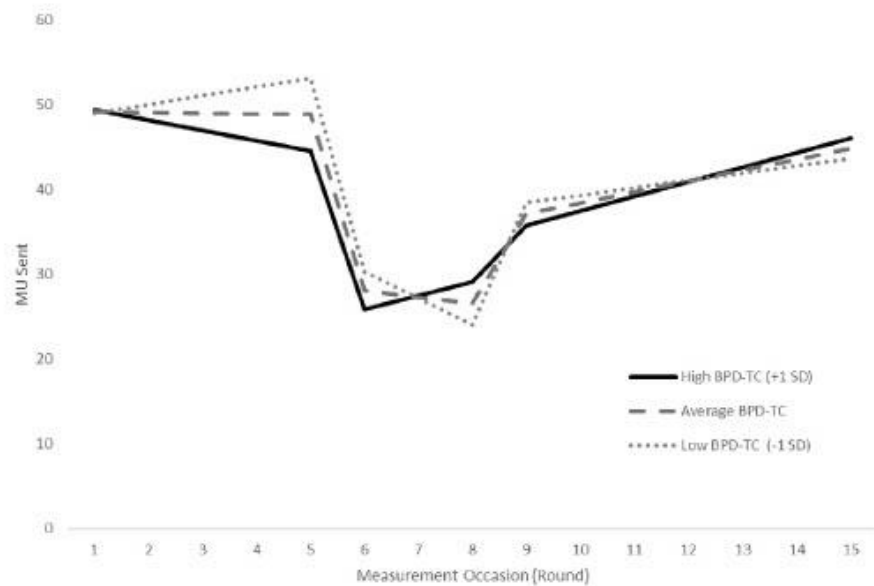


Fig 2. Number of MUs sent by the investor as a function of Borderline Personality Disorder Trait Count (BPD-TC). High BPD-TC are investors who scored one or more standard deviations above the mean number of BPD traits on the McLean Screening Instrument for Borderline Personality Disorder, and Low BPD-TC are investors who scored one or more standard deviations below the mean number of BPD traits.

<https://doi.org/10.1371/journal.pone.0236170.g002>

the relationship between BPD trait count and the trajectory of trust during the formation phase where we found a significant negative linear trend ($est_{PS \cdot BPD} = -1.12, p < .10$). This result indicates that individuals with a higher number of BPD traits have a more pronounced decline in MUs transferred during the trust formation phase. This was additionally confirmed in a model which isolated only the first 5 rounds of the game to measure trust change ($est_{PS \cdot BPD} = -1.02, p < 0.10$).

Dissolution transition. It was predicted that BPD would be associated with a larger immediate decrease in the level of trust immediately after the trust violation. The hypothesis for the relationship between BPD trait count and the dissolution transition was not supported. While there was a decrease in MUs transferred following the first exposure to the trust violation ($est_{DT} = -20.74, p < .001$), the size of the decrease appeared to be similar regardless of BPD trait count ($est_{DT \cdot BPD} = 2.13, n.s.$).

Dissolution slope. It was hypothesized that BPD would be associated with a slower rate of trust growth during the dissolution phase. As indicated by Table 5, while BPD trait count appeared to be a substantive moderator on the change in MUs transferred during the dissolution phase, the effect was not in the hypothesized direction. Rather, there was a significant positive linear trend between BPD trait count and the amount transferred during dissolution ($est_{DS \cdot BPD} = 2.33, p < .05$). This result indicates that individuals with a higher number of BPD traits have a more pronounced increase in the rate of MUs transferred during the dissolution phase, rather than a decline as predicted.

The comparative relative model is used to understand the impact of the trust violation on the intra-individual patterns of trust. It was hypothesized that BPD would be associated with a slower rate of trust growth during the dissolution phase relative to the trust growth observed in the formation phase. Once again, BPD was significantly associated with the rate at which

trust changed during dissolution relative to formation, but not in the hypothesized direction ($est_{DS^*BPD} = 3.45, p < .01$). In relative terms, BPD was associated with a rate of growth in MUs transferred during dissolution that was greater than the rate of growth during formation.

Restoration transition. It was predicted that BPD would be associated with a smaller increase in the level of trust immediately following the trust repair. As hypothesized, BPD trait count appeared to be a substantive moderator on the increase in the number of MUs transferred at the restoration transition ($est_{RT^*BPD} = -3.83, p < .05$). High BPD trait individuals had a significantly less pronounced increase in MUs transferred following the first instance of repair. This effect was even more pronounced when considering the shift in trust levels following repair relative to the level of trust predicted had the repair never occurred ($est_{RT^*BPD} = -6.16, p < .05$). In other words, for persons high on BPD trait count, trust levels increased immediately following restoration but to a significantly less extent than persons low on BPD trait count. Finally, when considering the shift in trust levels following repair relative to the formation phase, BPD was associated with a larger number of units transferred ($est_{RT^*BPD} = 7.43, p < .10$). In other words, the level of trust (at round 9) was greater than what would have been expected based on the formation phase pattern of trust change.

Restoration slope. It was predicted that BPD would be associated with a slower rate of trust growth during the restoration phase. Our hypothesis for the relationship between BPD trait count and the restoration slope was not supported. The linear pattern in MUs transferred during the restoration phase ($est_{RS} = 1.29, p < .01$) was at a similar rate regardless of BPD trait count ($est_{RS^*BPD} = .43, n.s.$).

With reference to relative change, it was predicted that BPD would be associated with a less pronounced rate of trust growth during the restoration phase relative to the formation phase. However, while BPD trait count was found to significantly moderate the rate of relative change during restoration, it was not in the hypothesized direction. BPD trait count was associated with a more pronounced positive linear trend in the number of MUs transferred during restoration relative to the linear trend in transfers observed in the formation phase ($est_{RS^*BPD} = 1.55, p < .05$). BPD was associated with a rate of growth in MUs transferred during restoration that was greater than the rate of growth during formation.

It was also predicted that BPD would be associated with a less pronounced rate of trust growth during the restoration phase even after taking into account the trajectory during dissolution. Results indicate a non-significant coefficient trending in the hypothesized direction ($est_{RS^*BPD} = -1.90, p = .13$). It should be noted however that we expected BPD trait count to be associated with faster dissolving trust in response to the violation, but we found the opposite; indeed, trust grew at a faster rate. In light of this, the non-significant trend described above suggests that, in comparison to the faster growth observed in the dissolution phase, trust growth was tempered during restoration at higher BPD trait counts.

Trust change as a function of BPD caseness. To support the clinical utility of our findings the level 2 analyses were repeated using the MSI-BPD recommended cutoff values to group participants into likelihood of "caseness" ($N = 16$) and "noncaseness" ($N = 218$). Results reported are limited to key findings, but a complete account can be found in the supporting files (S1 Table).

In contrast to the BPD trait model, BPD caseness was not found to modify the rate of trust growth during the formation phase, although there were non-significant trends in the expected direction ($est_{PS^*caseness} = -3.57, p = .12$). BPD caseness was found to positively moderate trust growth during the dissolution phase, both relative to the formation phase ($est_{DS^*caseness} = 14.32, p < .01$) and in absolute terms ($est_{DS^*caseness} = 10.74, p < .05$). In response to the repair, BPD caseness was associated with a smaller absolute increase in MUs transferred ($est_{RT^*caseness} = -14.08, p < .10$). For those meeting BPD caseness, trust levels were also lower after the first

instance of repair than that predicted by the trajectory of growth during the dissolution phase ($est_{RT}^{caseness} = -8.29, p < .10$). However, caseness was not found to modify the level of trust following repair relative to the level expected based on the formation trajectory ($est_{RT}^{caseness} = 21.32, p = .19$). Finally, the rate of trust growth during restoration for those meeting BPD caseness was faster than trust growth during formation ($est_{RS}^{caseness} = 6.02, p < .05$), and slower than trust growth during dissolution ($est_{RS}^{caseness} = -8.29, p < .01$).

Discussion

The present study used discontinuous growth modelling with a trust game to investigate the relationship between BPD and trust behaviours. It examined how BPD trait count modified the level and trajectory of trust as it formed with a new partner, dissolved in response to trust violation, and was rebuilt in response to trust repair. The results suggest that trust behaviour in individuals with a high number of BPD traits may be broadly classified as cautious and mistrustful in the beginning of a new relationship, even when the other party is behaving in a cooperative and trustworthy manner, and conversely, during and after a trust violation, trust appears to grow markedly, even in the face of repeated betrayals.

When interacting with a new and cooperative partner, high BPD trait individuals became progressively less trusting. That is, despite earning up to 32% profit on each investment made during this period, as the number of BPD traits increased, individuals progressively reduced the number of MUs transferred. A trend in this direction was also observed in those whose trait levels indicated likely caseness for BPD. This result augments previous findings indicating that average trust levels and trust growth are lower for people with BPD when playing the TG with a new and not uncooperative partner [17, 18]. Regarding the impact of BPD traits on trust behaviour following violation and repair, the findings highlight the benefits of using a method of analysis that is responsive to the dynamic, multiphasic nature of trust. While previous research found average trust levels during dissolution and restoration to be comparable between BPD patients and HC's [18], the current study found BPD had a paradoxical influence on trust patterns after violation and repair. If increasing mistrust defines the pre-violation trust patterns of high BPD trait count individuals, incessant trust growth would best characterise the violation and post-violation periods.

First, despite experiencing three consecutive rounds in which the trustee kept all or most of the investment, BPD trait count was associated with making progressively larger transfers, a finding contrary to what we expected based on King-Casas and colleagues observation that trustees with BPD were less likely to respond to diminishing cooperation via increasing reciprocity [16]. Given that the immediate reduction in the amount transferred following the first instance of violation was of a comparable magnitude across all levels of BPD symptomatology, even when the pre-violation trajectories were accounted for, it is likely that the three violation rounds were recognized as a norm violation irrespective of level of borderline pathology. This is consistent with findings that individuals with BPD can accurately appraise the fairness of trustee reciprocity [33]. It is possible that the decreases in investment size signalling diminishing trust were more nuanced in the King-Casas study [16], whereas our trust violation was unambiguous and therefore, may have prompted an unconventional response. The disparate coaxing behaviours between King-Casas' trustees and our investors may also reflect a property of the roles. It is possible that the powerlessness inherent in the trustee role—it is the investor who decides whether to initially engage—elicits a more aggressive response in people with BPD, while as investor, they can choose whether to be benevolent knowing that they can stop engaging at any point.

Regarding trust behaviours following repair, the size of the increase in MUs transferred following the initiation of trust repair became less pronounced as trait count increased. For persons who reported 9 or 10 BPD traits, there was a decrease rather than the expected increase. In other words, those with the highest levels of BPD symptoms paradoxically reduced the size of their next transfer in response to the trustee's return to cooperative play. It is possible that the first sign of reparative action by the trustee elicited caution and suspicion in higher trait individuals. In fact, if a prediction were to be made about how many MUs high BPD individuals would have sent at that point, had the violation continued, it would be a markedly higher amount than that sent in response to the repair. On the other hand, perhaps individuals with a higher BPD trait count had sought to coax higher returns from the trustee, and upon achieving this objective, stopped the coaxing behaviour and reduced their investments. Finally, it was found that that investors, in general, were on average observed to invest progressively larger amounts throughout the restoration phase. When this growth was considered relative to the growth observed in the formation phase, BPD traits were associated with a comparatively faster rate of growth. That is, trust in high BPD trait persons was observed to restore at a more accelerated rate than it was formed in the period before the violation took place. In fact, as trait count increased, trust grew at a faster rate during both dissolution and restoration than it had during formation.

This pattern of intraindividual fluctuations does not appear to map on to the intraindividual trust fluctuations of the general population [24, 26], with trust during each phase largely appearing to flow in the opposite direction. This suggests that although the ability to recognize norm violations does not appear to be compromised, BPD is associated with intraindividual changes in *trust behaviours* that are socially atypical. This has serious implications for individuals with BPD in terms of how they may be experienced by others during interpersonal interactions. Social cognition processes are believed to be engaged when individuals make strategic interpersonal decisions [60–62]. To maintain a mutually beneficial equilibrium in the TG, each party must be able to recognize, decipher, and respond appropriately to the signals sent. In the case of repeated interactions with the same partner such as in an iterated TG, the intraindividual fluctuations in behaviour communicate meaningful interpersonally relevant data.

In addition to being able to model the mind of the other, Kishida and colleagues [63] also propose that fair social exchange requires three computational capacities in each agent: to compute the social norms associated with such an exchange, to recognize deviations from said norms, and to respond appropriately considering these deviations. In other words, each agent must not only recognize and ascribe meaning to the intraindividual fluctuations of social signals emitted by their partner, but also determine if these fluctuations are socially normative. For example, previous research using a college student population found that the normative pattern in intraindividual fluctuations is that the *restoring* of trust appears to be a lengthier process than forming trust with a new partner [24]. The socially non-normative pattern of intraindividual fluctuations associated with BPD symptomatology may be perceived by others as incomprehensible, unexpected, and perhaps even odd interpersonal dynamics. In as few as 15 rounds of social exchange, the current study showed that borderline pathology predicts a paradoxical relational style that may invite confusion, whereby betrayal begets trust and cooperation, mistrust.

It is conceivable that these kinds of behaviours and preferences are likely to compromise the development of healthy relationships and may lead to relationship breakdowns or attract partners who may perpetuate these potentially deleterious relational dynamics [64]. Our results also support previous research in which individuals with BPD have been found to demonstrate greater acceptance of and perhaps a preference for inequitable treatment. For

instance, findings from economic game studies suggest that compared to HC's, BPD individuals are more likely to accept unfair offers [65] and reject fair offers [66], and express a greater preference for an unfair interaction partner, and lower preference for a fair interaction partner [67]. While much of the trust literature in borderline populations has focused on BPD tendencies towards mistrust and lack of cooperation, our analytical methodology has highlighted the other extreme, the tendency to engage in trusting behaviour in contexts warranting prudence. Indeed, findings from a social network analysis study has showed that BPD is associated with reduced discrimination in differentiating whom in their social network they seek advice and emotional support from [8]. That is, despite the tendency towards trust-compromising beliefs, appraisals, and behaviours, people with BPD may also trust more haphazardly or arbitrarily.

Limitations & future directions

The current study has several limitations. First, a clinical sample was not used. Previous research has primarily used patients with a BPD diagnosis, so it is possible our findings may not apply to individuals with severe and persistent BPD impairments. It is important to note however, that interpersonal disturbances in nonclinical samples of people with borderline traits are almost as profound as in clinical samples [68], and that young adults with sub-clinical borderline features are more likely to exhibit interpersonal dysfunction at a two-year follow up than their healthy counterparts [69]. Second, we used a simulated trustee rather than an actual human being, and therefore created interactions where the trustee was unconditionally cooperative or uncooperative, regardless of our participants' behaviour. While standardized trustee reciprocity was chosen intentionally to create the distinct phases of trust, doing so limits the conclusions that can be drawn since interpersonal trust is a dyadic process [70]. Third, participants were not offered financial incentives to participate. By not tying the game results to a financial reward, participants may have been less motivated to take the game seriously. Fourth, BPD is most typically associated with relational disturbances in close relationships. It is not clear whether the behaviour observed in a low stakes game with an anonymous partner would reflect trust behaviours in close personal relationships. For example, Miano et al. [71] found that within the context of intimate relationships, individuals with BPD compared to HC's appraised their romantic partner as less trustworthy after discussing a relationship threatening or personally threatening topic, whereas appraisal ratings were comparable following discussion of a neutral topic. It is therefore likely that the experience of a trust violation and repair within the context of an intimate relationship may evoke a more marked or varied response.

Our findings and methodology provide a solid foundation upon which researchers can examine the nuanced factors and processes that may underlie these incongruous trust dynamics, and therefore help inform more targeted interventions. For example, previous research has looked at the effect of underlying attachment insecurity on trust behaviours [e.g., 72–74]. Our methodology could be used to elicit potentially attachment salient events such as a trust violation, and examine whether attachment insecurity underpins the effect of borderline pathology on trust behaviours. Our findings should also be replicated using a patient sample to allow clinical inferences to be made.

Conclusions

To the best of our knowledge, this is the first study with a BPD focus to use discontinuous growth modelling with a trust game to examine trust as a dynamic and multiphasic process. Specifically, the study revealed the trajectories of trust as it forms, dissolves, and restores in response to trust violation and repair. It also explained how these trajectories varied as a function of the number of BPD traits reported. Showing how trust fluctuates within the individual

and how symptom count modifies the magnitude and direction of the fluxes, provides a richer and more nuanced understanding of how people with BPD traits engage with others in trust-altering contexts. This approach uncovered a paradoxical style of social exchange where social norms appear to be contradicted, thereby creating interpersonal encounters that are seemingly ambivalent, aberrant, and puzzling. By adopting a design and analytical methodology that recognizes the dynamic nature of trust, the study uniquely illustrated how relational disturbances may be produced and maintained in a BPD population.

Supporting information

S1 Dataset. Dataset for study.
(CSV)

S1 Table. Discontinuous mixed-effects growth models predicting trust as a function of BPD caseness after controlling for gender and cognitive reflective ability.
(DOCX)

Acknowledgments

The authors would like to thank Professors Paul Bliese and Jonas Lang for statistical consultation, and Yossef Abramov for developing the online trust game platform.

Author Contributions

Conceptualization: Gamze Abramov.

Data curation: Gamze Abramov.

Formal analysis: Gamze Abramov, Sebastien Miellet, Jason Kautz.

Investigation: Gamze Abramov.

Methodology: Gamze Abramov.

Project administration: Gamze Abramov.

Resources: Gamze Abramov.

Supervision: Sebastien Miellet, Brin F. S. Grenyer, Frank P. Deane.

Visualization: Gamze Abramov.

Writing – original draft: Gamze Abramov.

Writing – review & editing: Gamze Abramov, Sebastien Miellet, Jason Kautz, Brin F. S. Grenyer, Frank P. Deane.

References

1. Sansone RA, Sansone LA. Personality disorders: A nation-based perspective on prevalence. *Innov Clin Neurosci.* 2011; 8(4):13–8. PMID: 21637629.
2. Gunderson JG. Borderline personality disorder. *N Engl J Med.* 2011; 364(21):2037–42. <https://doi.org/10.1056/NEJMcp1007358> PMID: 21612472
3. Becker DF, Añez LM, Paris M, Grilo CM. Exploratory factor analysis of borderline personality disorder criteria in monolingual Hispanic outpatients with substance use disorders. *Psychiatry Res.* 2010; 178:305–8. <https://doi.org/10.1016/j.psychres.2009.03.016> PMID: 20472296
4. Sanislow CA, Grilo CM, McGlashan TH. Factor analysis of the DSM-III-R borderline personality disorder criteria in psychiatric inpatients. *The American Journal of Psychiatry.* 2000; 157(10):1629–33. <https://doi.org/10.1176/appi.ajp.157.10.1629> PMID: 11007717

5. Zanarini MC, Gunderson JG, Frankenburg FR, Chauncey DL. Discriminating borderline personality disorder from other axis II disorders. *The American Journal Of Psychiatry*. 1990; 147(2):161–7. <https://doi.org/10.1176/ajp.147.2.161> PMID: 2301653.
6. Gunderson JG, Stout RL, McGlashan TH, Shea MT, Morey LC, Grilo CM, et al. Ten-year course of borderline personality disorder: psychopathology and function from the Collaborative Longitudinal Personality Disorders study. *Arch Gen Psychiatry*. 2011; 68(8):827–37. <https://doi.org/10.1001/archgenpsychiatry.2011.37> PMID: 21464343.
7. Zanarini MC, Frankenburg FR, Reich DB, Fitzmaurice G. Time to attainment of recovery from borderline personality disorder and stability of recovery: A 10-year prospective follow-up study. *Am J Psychiatry*. 2010; 167(6):663–7. <https://doi.org/10.1176/appi.ajp.2009.09081130> PMID: 20395399.
8. Clifton A, Pilkonis PA, McCarty C. Social networks in borderline personality disorder. *Journal of Personality Disorders*. 2007; 21(4):434–41. <https://doi.org/10.1521/pedi.2007.21.4.434> PMID: 17685838
9. Lazarus SA, Cheavens JS. An examination of social network quality and composition in women with and without borderline personality disorder. *Personality Disorders-Theory Research and Treatment*. 2017; 8(4):340–8. <https://doi.org/10.1037/per0000201> PMID: 27347686
10. Lazarus SA, Southward MW, Cheavens JS. Do borderline personality disorder features and rejection sensitivity predict social network outcomes over time? *Pers Individ Dif*. 2016; 100:62–7. <https://doi.org/10.1016/j.paid.2016.02.032>
11. Zanarini MC, Frankenburg FR, Reich DB, Wedig MM, Conkey LC, Fitzmaurice GM. The course of marriage/sustained cohabitation and parenthood among borderline patients followed prospectively for 16 years. *Journal Of Personality Disorders*. 2015; 29(1):62–70. https://doi.org/10.1521/pedi_2014_28_147 PMID: 24963829.
12. 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. *Journal of Personality Disorders*. 2015; 29(5):575–609. <https://doi.org/10.1521/pedi.2015.29.5.575> PMID: 26393477
13. Barazandeh H. A systematic review of the relationship between early maladaptive schemas and borderline personality disorder/traits. *Pers Individ Dif*. 2016; 94:130–9. <https://doi.org/10.1016/j.paid.2016.01.021>
14. Bhar SS, Brown GK, Beck AT. Dysfunctional beliefs and psychopathology in borderline personality disorder. *Journal of Personality Disorders*. 2008; 22(2):165–77. <https://doi.org/10.1521/pedi.2008.22.2.165> PMID: 18419236
15. Barazandeh H, Kissane DW, Saeedi N, Gordon M. A systematic review of the relationship between early maladaptive schemas and borderline personality disorder/traits. *Pers Individ Dif*. 2016; 94:130–9. <https://doi.org/10.1016/j.paid.2016.01.021>
16. King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague RR. The rupture and repair of cooperation in borderline personality disorder. *Science*. 2008; 321(5890):806–10. <https://doi.org/10.1126/science.1156902> PMID: 18687957
17. Unoka Z, Seres I, Áspán N, Bódi N, Kéri S. Trust game reveals restricted interpersonal transactions in patients with borderline personality disorder. *Journal of Personality Disorders*. 2009; 23(4):399–409. <https://doi.org/10.1521/pedi.2009.23.4.399> PMID: 19663659
18. 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. <https://doi.org/10.1037/abn0000373> PMID: 30102052
19. Givertz M, Woszidlo A, Segrin C, Knutson K. Direct and indirect effects of attachment orientation on relationship quality and loneliness in married couples. *Journal of Social and Personal Relationships*. 2013; 30(8):1096–120. <https://doi.org/10.1177/0265407513482445>
20. Campbell L, Simpson JA, Boldry JG, Rubin H. Trust, variability in relationship evaluations, and relationship processes. *J Pers Soc Psychol*. 2010; 99(1):14–31. <https://doi.org/10.1037/a0019714> PMID: 20565183
21. Joel S, MacDonald G, Page-Gould E. Wanting to stay and wanting to go: Unpacking the content and structure of relationship stay/leave decision processes. *Soc Psychol Personal Sci*. 2018; 9(6):631–44. <https://doi.org/10.1177/1948550617722834>
22. Berg J, Dickhaut J, McCabe K. Trust, reciprocity, and social history. *Games and Economic Behavior*. 1995; 10(1):122. doi.org/10.1006/game.1995.1027.
23. Lewicki RJ, Tomlinson EC, Gillespie N. Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*. 2006; 32(6):991–1022. doi.org/10.1177/0149206306294405.

24. Korsgaard AM, Kautz J, Bliese PD, Samson K, Kostyszyn P. Conceptualising time as a level of analysis: New directions in the analysis of trust dynamics. *Journal of Trust Research*. 2018; 8(2):142–65. <https://doi.org/10.1080/21515581.2018.1516557>
25. Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: A cross-discipline view of trust. *Acad Manage Rev*. 1998; 23(3):393–404. <https://doi.org/10.5465/AMR.1998.926617>
26. Fulmer CA, Gelfand MJ. Trust after violations: Are collectivists more or less forgiving? *Journal of Trust Research*. 2015; 5(2):109. <https://doi.org/10.1080/21515581.2015.1051050>
27. Fulmer CA. *Dynamic trust processes after violation: Trust dissolution and restoration*. College Park: University of Maryland; 2010.
28. Fulmer CA, Gelfand MJ. How do I trust thee? Dynamic trust patterns and their individual and social contextual determinants. In: Sycara K, Gelfand M, Abbe A, editors. *Models for intercultural collaboration and negotiation*. Dordrecht: Springer Netherlands; 2013. p. 97–131.
29. Lount RB Jr., Zhong CB, Sivanathan N, Mumighan JK. Getting off on the wrong foot: The timing of a breach and the restoration of trust. *Pers Soc Psychol Bull*. 2008; 34(12):1601–12. doi.org/10.1177/0146167208324512. PMID: 19050335
30. Thayer A. *Trust trajectories as a function of violation type and repair efforts*. Electronic Theses and Dissertations: University of Central Florida; 2015.
31. Schilke O, Reimann M, Cook KS. Effect of relationship experience on trust recovery following a breach. *Proc Natl Acad Sci USA*. 2013; 110(38):15236–41. <https://doi.org/10.1073/pnas.1314857110> PMID: 24003151
32. Ebert A, Kolb M, Heller J, Edel MA, Roser P, Brüne M. Modulation of interpersonal trust in borderline personality disorder by intranasal oxytocin and childhood trauma. *Soc Neurosci*. 2013; 8(4):305–13. <https://doi.org/10.1080/17470919.2013.807301> PMID: 23802121
33. 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(1–2):224–33. <https://doi.org/10.1016/j.psychres.2010.11.012> PMID: 21129781
34. Zinchenko O, Arsalidou M. Brain responses to social norms: Meta-analyses of fMRI studies. *Hum Brain Mapp*. 2018; 39(2):955–70. <https://doi.org/10.1002/hbm.23895> PMID: 29160930
35. Rilling JK, Sanfey AG. The neuroscience of social decision-making. *Annu Rev Psychol*. 2011; 62:23–48. <https://doi.org/10.1146/annurev.psych.121208.131647> PMID: 20822437
36. Bliese PD, Lang JWB. Understanding relative and absolute change in discontinuous growth models: Coding alternatives and implications for hypothesis testing. *Organizational Research Methods*. 2016; 19(4):562–92. <https://doi.org/10.1177/1094428116633502>
37. Singer JD, Willett JB. *Applied longitudinal data analysis: Modeling change and event occurrence* Oxford University Press; 2003.
38. Bryk AS, Raudenbush SW. *Hierarchical linear models: Applications and data analysis methods* Newbury Park, CA: Sage 2002.
39. Frederick S. Cognitive reflection and decision making. *The Journal of Economic Perspectives*. 2005; 19(4):25. <https://doi.org/10.1257/089533005775196732>
40. Corgnet B, Espín AM, Hemán-González R, Kujal P, Rässenti S. To trust, or not to trust: Cognitive reflection in trust games. *Journal of Behavioral and Experimental Economics*. 2016; 64:20–7. <https://doi.org/10.1016/j.socec.2015.09.008>
41. Ruocco AC. The neuropsychology of borderline personality disorder: A meta-analysis and review. *Psychiatry Res*. 2005; 137(3):191–202. <https://doi.org/10.1016/j.psychres.2005.07.004> PMID: 16297985
42. McClure G, Hawes DJ, Dadds MR. Borderline personality disorder and neuropsychological measures of executive function: A systematic review. *Personality and Mental Health*. 2016; 10(1):43–57. <https://doi.org/10.1002/pmh.1320> PMID: 26381859
43. Paret C, Jermen-Steinmetz C, Schmah I. Disadvantageous decision-making in borderline personality disorder: Partial support from a meta-analytic review. *Neurosci Biobehav Rev*. 2017; 72:301–9. <https://doi.org/10.1016/j.neubiorev.2016.11.019> PMID: 27914943
44. Lazarus SA, Cheavens JS, Festa F, Zachary Rosenthal M. Interpersonal functioning in borderline personality disorder: A systematic review of behavioral and laboratory-based assessments. *Clin Psychol Rev*. 2014; 34(3):193–205. <https://doi.org/10.1016/j.cpr.2014.01.007> PMID: 24534643
45. Berenson KR, Gregory WE, Glaser E, Romirowsky A, Rafaeli E, Yang X, et al. Impulsivity, Rejection Sensitivity, and Reactions to Stressors in Borderline Personality Disorder. *Cognit Ther Res*. 2016; 40(4):510–21. <https://doi.org/10.1007/s10608-015-9752-y> PMID: 27616800
46. Johnson ND, Mislin AA. Trust games: A meta-analysis. *J Econ Psychol*. 2011; 32(5):865–89. <https://doi.org/10.1016/j.joep.2011.05.007>

47. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. Washington, DC: Author; 2013.
48. Sitkin SB, Roth NL. Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science*. 1993; 4(3):367. doi.org/10.1287/orsc.4.3.367.
49. Murnighan JK. Defectors, vulnerability, and relative power: Some causes and effects of leaving a stable coalition. *Human Relations*. 1981; 34(7):589. https://doi.org/10.1177/001872678103400706
50. Zanarini MC, Vujanovic AA, Parachini EA, Boulanger JL, Frankenburg FR, Hennen J. A screening measure for BPD: The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). *Journal of Personality Disorders*. 2003; 17:568–73. https://doi.org/10.1521/pedi.17.6.568.25355 PMID: 14744082
51. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th-revised ed. Washington, DC: Author; 2000.
52. Berwick DM, Murphy JM, Goldman PA, Ware J.E. Jr., Barsky AJ, Weinstein MC. Performance of a Five-Item Mental Health Screening Test. *Med Care*. 1991; 29(2):169.
53. Gemans S, Van Heck GL, Moran P, Hodiamont PPG. The Self-report Standardized Assessment of Personality-abbreviated Scale: Preliminary results of a brief screening test for personality disorders. *Personality and Mental Health*. 2008; 2(2):70–6. https://doi.org/10.1002/pmh.34
54. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. nlme: Linear and nonlinear mixed effects models. 2019.
55. R Development Core Team. R: A Language and Environment for Statistical Computing. 3.6.1 ed. Vienna: R Foundation for Statistical Computing; 2018.
56. Snijders TAB, Bosker RJ. Multilevel analysis: an introduction to basic and advanced multilevel modeling / Snijders Tom A. B. and Bosker Roel J.: SAGE; 1999.
57. Yeo GB, Neal A. A multilevel analysis of effort, practice, and performance: Effects of ability, conscientiousness, and goal orientation. *J Appl Psychol*. 2004; 89(2):231–47. https://doi.org/10.1037/0021-9010.89.2.231 PMID: 15065972
58. Niessen C, Jimmieson NL. Threat of resource loss: The role of self-regulation in adaptive task performance. *J Appl Psychol*. 2016; 101(3):450–62. https://doi.org/10.1037/apl0000049 PMID: 26348477
59. Bliese PD, Ployhart RE. Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*. 2002; 5(4):362–87. https://doi.org/10.1177/109442802237116
60. Frith CD, Singer T. Review. The role of social cognition in decision making. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2008; 363(1511):3875–86.
61. Hampton AN, Bossaerts P, O'Doherty JP. Neural correlates of mentalizing-related computations during strategic interactions in humans. 2008:6741.
62. Lee VK, Harris LT. How social cognition can inform social decision making. *Front Neurosci*. 2013; 7:13.
63. Kishida KT, King-Casas B, Montague PR. Neuroeconomic approaches to mental disorders. *Neuron*. 2010; 67(4):543–54. https://doi.org/10.1016/j.neuron.2010.07.021 PMID: 20797532
64. Bouchard S, Sabourin S, Lussier Y, Villeneuve E. Relationship quality and stability in couples when one partner suffers from borderline personality disorder. *J Marital Fam Ther*. 2009; 35(4):446–55. https://doi.org/10.1111/j.1752-0606.2009.00151.x PMID: 19785701
65. Polgar P, Fogd D, Unoka Z, Siraly E, Csukly G. Altered social decision making in borderline personality disorder: An ultimatum game study. *Journal of Personality Disorders*. 2014; 28(6):841–52. https://doi.org/10.1521/pedi.2014.28.142 PMID: 24963831
66. De Panfilis C, Schito G, Generali I, Ossola P, Marchesi C, Grecucci A, et al. Emotions at the border: Increased punishment behavior during fair interpersonal exchanges in borderline personality disorder. *J Abnorm Psychol*. 2019; 128(2):162–72. https://doi.org/10.1037/abn0000404 PMID: 30714797
67. Jeung H, Vollmann M, Herpertz SC, Schwieren C. Consider others better than yourself: Social decision-making and partner preference in borderline personality disorder. *J Behav Ther Exp Psychiatry*. 2018. https://doi.org/10.1016/j.jbtep.2018.11.004 PMID: 30458960
68. Tolpin LH, Gunthert KC, Cohen LH, O'Neill SC. Borderline personality features and instability of daily negative affect and self-esteem. *J Pers*. 2004; 72(1):111–37. https://doi.org/10.1111/j.0022-3506.2004.00258.x PMID: 14686886.
69. Trull TJ, Useda D, Conforti K, Doan B-T. Borderline personality disorder features in nonclinical young adults: 2 Two-year outcome. *J Abnorm Psychol*. 1997; 106(2):307–14. https://doi.org/10.1037//0021-843x.106.2.307
70. Simpson JA. Foundations of interpersonal trust. In: Kruglanski AW, Higgins ET, editors. *Social psychology: Handbook of basic principles*. 2nd ed. New York: Guilford; 2007. p. 587–607.

71. Miano A, Fertuck EA, Roepke S, Dziobek I. Romantic relationship dysfunction in borderline personality disorder—a naturalistic approach to trustworthiness perception. *Personality Disorders: Theory, Research, and Treatment*. 2017; 8(3):281–6. <https://doi.org/10.1037/per0000196>
72. McClure MJ, Bartz JA, Lydon JE. Uncovering and overcoming ambivalence: The role of chronic and contextually activated attachment in two-person social dilemmas. *J Pers*. 2013; 81(1):103–17. <https://doi.org/10.1111/j.1467-6494.2012.00788.x> PMID: 22432960.
73. 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(5):556–63. <https://doi.org/10.1093/scan/nsq085> PMID: 21115541
74. Taheri M, Rotshtein P, Beierholm U. The effect of attachment and environmental manipulations on cooperative behavior in the prisoner's dilemma game. *PLoS One*. 2018; 13(11):e0205730. <https://doi.org/10.1371/journal.pone.0205730> PMID: 30419027

Appendix B: Published Manuscript: The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count (Chapter 3)

Abramov, G., Kautz, J., Miellel, S., & Deane, F. P. (2022). The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count. *Journal of Psychopathology and Behavioral Assessment*. <https://doi.org/10.1007/s10862-022-09965-9>



The Influence of Attachment Style, Self-protective Beliefs, and Feelings of Rejection on the Decline and Growth of Trust as a Function of Borderline Personality Disorder Trait Count

Gamze Abramov¹ · Jason Kautz² · Sebastien Miellet¹ · Frank P. Deane¹

Accepted: 7 February 2022
© The Author(s) 2022

Abstract

Borderline personality disorder (BPD) is associated with paradoxical trust behaviours, specifically a faster rate of trust growth in the face of trust violations. The current study set out to understand whether attachment style, self-protective beliefs, and feelings of rejection underpin this pattern. Young adults ($N=234$) played a 15-round trust game in which partner cooperation was varied to create three phases of trust: formation, dissolution, and restoration. Discontinuous growth modelling was employed to observe whether the effect of BPD trait count on trust levels and growth is moderated by fearful or preoccupied attachment style, self-protective beliefs, and feelings of rejection. Results suggest that the slower rate of trust formation associated with BPD trait count was accounted for by feelings of rejection or self-protective beliefs, both of which predicted a slower rate of trust growth. The faster rate of trust growth in response to trust violations associated with BPD trait count was no longer significant after self-protective beliefs were accounted for. Interventions targeting self-protective beliefs and feelings of rejection may address the trust-based interpersonal difficulties associated with BPD.

Keywords Borderline personality disorder · Trust · Cooperation · Attachment style · Rejection · Social cognition

Relational disturbances are a hallmark feature of borderline personality disorder (BPD) and are present among subclinical populations (Tolpin et al., 2004), persisting even after symptoms remit (Zanarini et al., 2010). Individuals with BPD tend to have intense relationships marred by conflict, instability, and rupture (Bouchard et al., 2009; Clifton et al., 2007; Lazarus et al., 2020). Aberrant trust processes have been proposed as a contributor to impaired interpersonal functioning in BPD (Poggi et al., 2019). Trust behaviours have been examined using behavioural economic experimental paradigms such as the trust game (Berg et al., 1995). Evidence suggests that in these trust game interpersonal exchanges, BPD is associated with reduced cooperation (King-Casas et al., 2008) and increased mistrust (Unoka et

al., 2009), even when engaging with a cooperative partner (Abramov et al., 2020; Liebke et al., 2018).

Currently, very little is known about the mechanisms that underpin the trust practices of individuals with BPD. Individuals with BPD typically develop mistrustful models of others, endorsing beliefs that others will betray, exploit, and deceive (Barazandeh et al., 2016; Bhar et al., 2008). As such, trust behaviours among individuals with BPD may be modulated by these beliefs. There is also growing evidence that situations or states which typically promote prosocial behaviours appear to have a paradoxical effect among individuals with borderline pathology (Abramov et al., 2020; Ebert et al., 2013; Liebke et al., 2018). For example, individuals with BPD appear to engage in less trusting behaviours following social acceptance than following social rejection (Liebke et al., 2018), suggesting rejection experiences may play a mechanistic role in trust processes in BPD. More recently, evidence suggests individuals who endorse many BPD symptoms appear to withhold trust when engaging with a partner whose actions appear trustworthy and cooperative, yet make more trusting gestures when that partner becomes uncooperative (Abramov et al., 2020). A similarly

✉ Gamze Abramov
ga385@uowmail.edu.au

¹ School of Psychology, University of Wollongong, Wollongong, NSW, Australia

² Naveen Jindal School of Management, University of Texas, Dallas, Texas, United States of America

paradoxical effect was observed among individuals with BPD who compared to controls, appeared to become less trusting following administration of the neuropeptide oxytocin, known for its prosocial effects on behaviour and used by researchers to activate the attachment system (Ebert et al., 2013), highlighting the potential role of attachment disturbance in modulating trust behaviours among individuals with BPD. The current study aims to understand the mechanisms underpinning the paradoxical trust behaviours associated with BPD by examining whether attachment style, self-protective beliefs, and feelings of rejection moderate the effects of BPD trait count on trust behaviours in response to cooperative and uncooperative partner play in a social exchange.

A brief overview of the trust game will be provided followed by a detailed rationale for the relationship between the study variables, BPD, and trust. In the trust game, trust is operationalized as the proportion of a monetary endowment (monetary units, MU) an individual, known as the *trustor*, is willing to entrust to an anonymous party, the *trustee* (Berg et al., 1995). The amount sent is multiplied by a factor before being received by the trustee who then has the option of returning a proportion to the trustor, which may result in a loss or profit to the trustor. Previous research has used the trust game to examine trust as a dynamic phenomenon whereby trustee reciprocity (trustworthiness) has been modified to create phases of trust (e.g., Abramov et al., 2020; Audrey Korsgaard et al., 2018; Fulmer & Gelfand, 2013; Liebke et al., 2018). Specifically, these are *trust formation* (initial stage of trust game where trustee reciprocity is held at levels that incur no loss and potential gain to the trustor), *trust dissolution* (trustee reciprocity results in loss to the trustor, constituting a violation of trust), and *trust restoration* (trustee reciprocity returns to pre-violation levels, constituting gestures of trust reparation).

Attachment Style and Trust

The socially atypical vicissitudes in trust associated with borderline pathology can be understood through the lens of attachment theory (Bowlby, 1969, 1973, 1980). Aetiological models of BPD suggest it is a disorder that in part stems from attachment disturbances arising from suboptimal, adverse, or invalidating caregiver experiences (Fonagy et al., 1995; Gunderson & Lyons-Ruth, 2008; Linehan, 1993). Interpersonal trust involves the belief that a person cares about one's needs and can be depended upon (Rempel et al., 1985) and securely attached individuals believe that significant others will be available, caring, and responsive, leading Mikulincer (1998) to conclude that trust is an integral tenet of secure attachment. The adult attachment styles associated

with BPD are fearful-avoidant attachment, and to a lesser extent, anxious-preoccupied attachment (Agrawal et al., 2004). Both styles endorse a negative model of the self in terms of personal insecurity, but those endorsing a more preoccupied style seek intimacy and connection while those endorsing a fearful attachment style avoid or find it difficult to become close to and dependent on others (Bartholomew & Horowitz, 1991). Research looking at trust and attachment style has used a dual-dimensional conceptualization of adult attachment. The dimension of attachment-related anxiety, refers to the degree to which an individual is worried about being rejected by the other, and attachment-related avoidance, refers to the degree to which strategies are used to downregulate attachment needs in relational contexts, with high scores indicating a discomfort with being close or dependent on the other (Fraley et al., 2015). Both fearful and preoccupied attachment styles reflect increased attachment-related anxiety, but only the fearful attachment style also has elevated levels of attachment-related avoidance.

The findings on the effect of attachment insecurity on trust behaviours are varied, and almost exclusively studied in non-clinical populations. Anxiously attached individuals were found to be more hesitant and more mistrustfully inconsistent in their responses in a social dilemma game (M. J. McClure et al., 2013). Likewise, under conditions of uncertainty where participants could randomly gain or lose money, participants became more cooperative, but this effect was muted among individuals with high levels of attachment anxiety or avoidance (Taheri et al., 2018). In contrast, there is some evidence that trust formation is positively associated with attachment anxiety (Fett et al., 2016). Furthermore, individuals with high levels of attachment anxiety may be more willing to self-sacrifice in favour of others. For example, in a bargaining game, anxiously attached individuals were both more generous in how much they were willing to offer the other player and more willing to accept less generous offers by the other player, in comparison to individuals with an avoidant attachment, who offered less and trended towards rejecting more offers (Almakias & Weiss, 2012). The authors suggested that anxiously attached persons are so concerned with gaining acceptance and avoiding abandonment, that they will forego monetary gain. On the other hand, they reasoned that avoidant individuals seek to avoid situations which might activate their attachment system, such as being exploited or unfairly rejected by the other. By making small offers, avoidant individuals can reason that any rejection that ensues was due to the low offer rather than a personal rejection (Almakias & Weiss, 2012). Curiously, this pattern observed in avoidantly attached individuals was found to reverse after administration of the neuropeptide oxytocin (De Dreu, 2012). However, the effects of oxytocin on those with a fearful attachment style

appears to be detrimental to cooperation. In a mixed sample of patients with BPD and healthy controls who played a game that incentivised mutual cooperation, those with a preoccupied attachment style became more cooperative following administration of oxytocin, while individuals with a fearful attachment style became less cooperative (Bartz et al., 2011). This finding was replicated by Ebert (2013) who found a similar effect among individuals with BPD. This suggests that a fearful attachment style may promote distrust and reduced willingness to cooperate under conditions of attachment arousal.

The current study examines whether attachment style underpins trust behaviours associated with BPD, and if so, whether there is a different effect based on the nature of the attachment disturbance. The extant research suggests that attachment anxiety is associated with heightened ambivalence about trusting but greater willingness to self-sacrifice in interpersonal transactions. Attachment avoidance appears to be associated with increased mistrust and a greater likelihood of behaving in a self-protective manner. Activation of attachment arousal also appears to have an adverse effect on trust behaviours for fearfully attached individuals, who report high levels of both attachment anxiety and avoidance. Endorsement of a fearful attachment style is expected to amplify the negative relationship between BPD and trust, thereby promoting lower levels of trust, while endorsement of a preoccupied attachment style is expected to mitigate that relationship, promoting higher levels of trust.

Self-protective Beliefs and Trust

Individuals with BPD endorse the belief that others will deceive, betray, and exploit (Barazandeh et al., 2016; Bhar et al., 2008). While none of the studies in the BPD literature have specifically examined the effect of holding such a priori beliefs on behavioural trust, there is evidence that in the absence of feedback regarding actual partner trustworthiness during a trust game, BPD is associated with greater mistrust (Unoka et al., 2009). Dubbing these findings as representative of 'unbiased trust,' Unoka and colleagues (2009) also found that patients with BPD predict a less favourable outcome of the game than controls, suggesting that their comparatively mistrustful behaviour and predictions may reflect a disposition towards perceiving others as untrustworthy. The current study will examine the effect of the a priori belief that one must protect oneself from betrayal by others on the relationship between BPD and trust behaviours. It is anticipated that self-protective beliefs will amplify the negative relationship between BPD and trust behaviours.

Feelings of Rejection and Trust

Interpersonal trust may also be influenced by negative affect that is present before the interpersonal exchange or elicited during the exchange. Events related to social rejection or abandonment in particular are known precipitants of emotional distress in BPD (Chapman et al., 2015; Staebler et al., 2011; Stiglmayr et al., 2005). Increased negative emotional arousal has been demonstrated to modify social processing in BPD (Dziobek et al., 2011; Wolff et al., 2007). Poggi and colleagues (2019) proposed that the mistrustful appraisals that are believed to underpin relational disturbances in BPD, may do so in conjunction with rejection sensitivity, another known mechanism of atypical social processing in BPD populations (Foxhall et al., 2019). Rejection sensitivity refers to a cognitive-affective processing disposition in which inevitable rejection by others is anxiously presumed, readily perceived, and overreacted to, often with hostility, attempts to control the other, or withdrawal (Downey & Feldman, 1996). Individuals high in rejection sensitivity expect that others will reject them and approach relationships with hypervigilance and hypersensitivity to signs of potential rejection, responding to actual or perceived rejection in ways that may compromise the relationship.

Rejection sensitivity has been found to mediate the relationship between BPD traits and an untrustworthiness bias for appraisals of neutral unfamiliar faces (Miano et al., 2013; Richetin et al., 2018). However, when it comes to trust behaviour, rejection and acceptance appear to have a paradoxical effect for individuals with BPD. For example, Liebke et al. (2018) primed participants with a social activity in which they experienced either social acceptance or social rejection. Among participants with BPD, those who had been primed with acceptance feedback invested significantly less than those primed with feedback of rejection. The more positive the feedback was, relative to what was expected, the smaller the corresponding investment. The authors suggested that receiving feedback of social acceptance, particularly when one expected to be rejected, triggered defences in those with BPD resulting in more withholding behaviour (Liebke et al., 2018). The current study will examine the impact pre-existing feelings of rejection on the relationship between BPD and trust behaviours. It is predicted that baseline feelings of rejection moderate the relationship between BPD and trust behaviours by increasing mistrustful behaviours during the cooperative phases of the trust game and increasing trustful behaviours during the uncooperative phases of the trust game.

Finally, there is evidence suggesting that females exhibit less trusting behaviours in the trust game (Johnson & Mislin, 2011). Given the gender bias associated with BPD (American Psychiatric Association, 2013), the main effects

of gender will be controlled. Cognitive reflective ability, that is, the ability to engage in conscious deliberation rather than respond impulsively (Frederick, 2005), has also been associated with a greater propensity to trust in the trust game (Corgnet et al., 2016). As BPD has been associated with deficits in executive functioning impairments (see G. McClure et al., 2016; Ruocco, 2005), social problem-solving (see Lazarus et al., 2014), altered decision making (see Paret et al., 2017), and increased impulsivity in interpersonal contexts (Berenson et al., 2016), the main effects of cognitive reflective ability will be controlled.

Method

The study was approved by the University of Wollongong ethics committee (HE2017/253). All participants provided informed consent. These data are part of a larger research project from which there has been one publication (Abramov et al., 2020). The current study is focused on data not previously published.

Participants

Participants ($N = 234$; 64% female; $M = 20.87$ years, $SD = 5.66$ years) were undergraduate students from a large Australian university who were invited to take part in an online study looking at the relationship between economic decision making and personality variables in exchange for course credit.

The Trust Game

The current study used a 15-round version of the trust game (Berg et al., 1995) with trust operationalized as the proportion of allocated monetary units (MU) sent to a trustee for investment (Ben-Ner & Halldorsson, 2010). At the start of each of round the trustor was allocated 100 MU by the experimenter and given the option to send the trustee any proportion from 0 to 100% for investment. The amount sent was automatically tripled, and the trustee could repay any proportion from 0 to 100% of the tripled investment to the trustor.

Procedure

Participants were all assigned the role of trustor and were paired with another anonymous participant who was, in fact, a computer program. Trustee repayments were programmed so that following rounds 1-4 and 8-14, repayments were randomized to fall between 34 and 44% of the tripled investment, providing the trustor with a return the equivalent of

the original investment plus up to 32% profit. Following rounds 5-7, trustee repayments were randomized to fall between 0 and 10% of the tripled investment, providing the trustor with a return equivalent to losing 70–100% of their original investment¹. This loss was designed to simulate a trust violation. Based on this repayment schedule, trustor investments can be divided into three distinct trust phases: formation (rounds 1-5), dissolution (rounds 6-8), and restoration (rounds 9-15). Overall, the average MU invested per round for each of the trust phases was 46.16 ($SD = 18.95$) during formation, 24.54 ($SD = 19.14$) during dissolution, and 38.4 ($SD = 23.29$) during restoration.

Measures

Trust

The number of MU's (0-100) that participants sent to the trustee in each round, represents a single behavioural measure of trust. Each participant provided 15 trust measures in total.

BPD

The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD; Zanarini et al., 2003) was used to assess DSM-5 BPD trait count. The MSI-BPD is a 10-item screening instrument for BPD, with very good internal consistency (Cronbach's $\alpha = 0.84$, $N = 234$). BPD in the current study is operationalized as the number of MSI-BPD items endorsed (0-10), rather than a clinical diagnosis. The current sample endorsed a median of one BPD trait ($M = 1.63$, $SD = 2.34$). While MSI-BPD requires endorsement of 7 out of 10 items for clinical diagnoses of BPD, endorsement of three or more items is reflective of sub-clinical borderline pathology (CLPS: Gunderson et al., 2011).

Attachment Style

The Relationship Questionnaire (RQ; Bartholomew & Horowitz, 1991) is a 4-item questionnaire designed to measure adult attachment style. The RQ consists of four paragraphs, each describing an attitude toward relationships as representative of one of four attachment styles (secure, dismissing, preoccupied, and fearful). For the current study only the preoccupied and fearful attachment styles were reported. Fearful attachment was characterized as: "I am uncomfortable getting close to others. I want emotionally close relationships, but I find it difficult to trust others completely, or to depend on them. I worry that I will be hurt

¹ Due to rounding to the nearest whole number, investments of \$1 did not incur a loss.

if I allow myself to become too close to others.” Preoccupied attachment was characterized as: “I want to be completely emotionally intimate with others, but I often find that others are reluctant to get as close as I would like. I am uncomfortable being without close relationships, but I sometimes worry that others don’t value me as much as I value them.” Participants were asked to rate each paragraph on an 11-point Likert-type scale of 0 (not at all like me) to 10 (very much like me).

Self-protective Beliefs

Self-protective beliefs were assessed using the Personality Beliefs Questionnaire—Borderline Personality Subscale (PBQ-BPD; Butler et al., 2002), a 14-item subset of the Personality Beliefs Questionnaire (PBQ; Beck & Beck, 1991). The PBQ-BPD subscale was developed on the basis of PBQ items that discriminated 84 BPD patients from 204 patients with other personality disorders (Bhar et al., 2008; Butler et al., 2002). Participants were asked to endorse each of the 14 items on a 5-point Likert-type scale from 0 (“I don’t believe it at all”) to 4 (“I believe it totally”). The subscale has demonstrated adequate internal consistency and discriminant validity (Butler et al., 2002). A study examining the factor structure of the PBQ-BPD using exploratory factor analysis found three factors: dependency, distrust, and the belief that one should take preemptive action to avoid threat (Bhar et al., 2008). In the current study, the three items comprising this preemptive action factor were used to provide a measure of self-protective beliefs (e.g. “People will get me if

I don’t get them first”). Cronbach’s alpha was $\alpha = .74$, $N = 234$.

Feelings of rejection

Feelings of rejection were assessed using four items used previously in a study that assessed affective states in patients with BPD (Gadassi et al., 2014). Before playing the game participants were prompted to rate on a 5-point Likert-type scale (0 = not at all, 4 = extremely) the extent to which they were currently experiencing specific emotions related to rejection (e.g. “At the present moment I feel rejected by others”). In the current study, internal consistency for pre-game rejection (Cronbach’s $\alpha = 0.74$, $N = 234$) was adequate, but lower than the 0.91 found in a sample of BPD patients (Gadassi et al., 2014).

Cognitive Reflective Ability

Cognitive reflective ability was measured using the Cognitive Reflection Test (CRT; Frederick, 2005), a three-item measure of the willingness to engage in deliberation during a cognitive task. Each item is a deceptively simple mathematical problem in which an intuitive/impulsive yet incorrect answer must be suppressed in order to calculate the correct response. The number of items answered correctly were summed to provide a CRT score ranging from 0 to 3. The overall sample mean of 1.03 was consistent with previous findings (see Frederick, 2005).

Table 1 Coding and interpretation of change variables in the discontinuous growth model

Round (R)	TIME	Dissolution Transition (DT)	Dissolution Slope (DS)	Restoration Transition (RT)	Restoration Slope (RS)
Measurement occasion in the trust game	Linear change of MUs transferred in the formation phase (R1 to R5)	Difference in level of MUs transferred immediately following the trust violation (R6 vs. R5)	Linear change of MUs transferred in the dissolution phase (R6 to R8)	Difference in level of MUs transferred immediately following the trust repair (R9 vs. R8)	Linear change of MUs transferred in the restoration phase (R9 to R15)
1	0	0	0	0	0
2	1	0	0	0	0
3	2	0	0	0	0
4	3	0	0	0	0
5	4	0	0	0	0
6	4	1	0	0	0
7	4	1	1	0	0
8	4	1	2	0	0
9	4	1	2	1	0
10	4	1	2	1	1
11	4	1	2	1	2
12	4	1	2	1	3
13	4	1	2	1	4
14	4	1	2	1	5
15	4	1	2	1	6

Data Analyses

The NLME package (Pinheiro et al., 2019) included in the open source software R (R Development Core Team, 2018) was used to conduct mixed-effect discontinuous growth modelling (DGM) analyses (Bliese & Lang, 2016; Singer & Willett, 2003). These analyses assessed investment occasions (rounds) at Level 1 nested within individuals at Level 2.

Level 1 parameters were coded based on the framework recommended by Singer and Willett (2003) and Bliese and Lang (2016) to create a matrix of time covariates that examine change in the average level of trust between each of the trust phases (i.e., formation, dissolution, and restoration) along with the growth of trust within each phase. To examine how individuals respond first to a trust violation and then to a trust restoration, change variables were such that the transition coefficients reflected the previous stage as the baseline for interpretation while the growth coefficients reflected growth relative to nil growth. *Time* was coded to capture the linear growth in trust during the formation phase. *Dissolution transition* (DT) and *Dissolution slope* (DS) were coded to represent the change in the average level of trust moving from the formation phase to the dissolution phase and the rate of trust growth during the dissolution phase, respectively. *Restoration transition* (RT) and *Restoration slope* (RS) were coded to represent the change in the average level of trust moving from the dissolution phase to the restoration phase and rate of trust growth during the restoration phase, respectively. The full DGM time-covariate matrix is presented in Table 1. Due to space constraints, the authors wish to encourage those interested in further understanding of the covariate matrix to read Bliese and Lang (2016) and/or Bliese, Kautz, and Lang (2020).

Initially Level 1 change was examined by including only Level 1 predictors into the DGM. Next, the influence of between-individual factors was examined by including BPD and one of the other Level 2 predictors (preoccupied attachment, fearful attachment, self-protection beliefs, feelings of rejection) to account for differences in Level 1 change, while also controlling for the main effects of gender and cognitive ability. All tests conducted were two-tailed, and a criterion level of $p < .10$ was used for all cross-level interactions effects to account for insufficient power to detect cross-level interactions as a result of reduced parameter reliability in multilevel analysis (Snijders & Bosker, 1999).

Results

Descriptive Data and Intercorrelations

Table 2 presents the means, standard deviations, and intercorrelations for BPD, attachment styles, self-protective beliefs, and pre-game feelings of rejection. BPD had a medium positive association with fearful attachment style, self-protective beliefs, and pre-game feelings of rejection, and a small positive association with preoccupied attachment style. Trust formation has a small negative association with BPD. None of the other trust phases is significantly associated with any of the predictors.

To understand the mechanisms underpinning the trust behaviour patterns associated with BPD, a series of 4 models were created which, in addition to including BPD as a predictor, also examined the effects of individuals differences in preoccupied and fearful attachment styles, self-protective beliefs, and feelings of rejection (see Table 3). To be consistent with previous research (Abramov et al., 2020), the main effects of gender and cognitive reflective ability were also

Table 2 Intercorrelations, Means, and Standard Deviations of Study Variables

	1	2	3	4	5	6	7	8
1 BPD	(0.84)							
2 Preoccupied Attachment	0.25	-						
3 Fearful Attachment	0.34	0.35	-					
4 Self-Protective Beliefs	0.38	0.26	0.31	(0.74)				
5 Feelings of Rejection	0.34	0.26	0.24	0.38	(0.74)			
6 Trust _{Formation}	-0.18	0.03	-0.08	-0.11	-0.12	-		
7 Trust _{Dissolution}	-0.01	0.06	-0.07	0.06	0.00	0.34	-	
8 Trust _{Restoration}	-0.08	0.11	-0.04	0.07	-0.02	0.44	0.55	-
Mean	1.63	4.20	4.56	1.03	2.25	46.16	24.54	38.40
SD	2.34	2.70	2.93	0.88	0.80	18.95	19.14	23.29
N	234	233	233	234	234	234	234	234

Trust_{Formation} reflects average trust during the formation phase. Trust_{Dissolution} reflects average trust during the dissolution phase. Trust_{Restoration} reflects average trust during the restoration phase

Spearman correlations are reported in the lower half. Alphas are reported on the diagonal. Values equal to or above |0.22| are significant at $p < .01$ level. Values equal to or above |0.18| are significant at $p < .05$ level. Values equal to or above |0.16| are significant at $p < .10$ level

Table 3 Discontinuous Mixed-Effects Growth Models Predicting Trust as a Function of BPD Trait Count (BPD), Attachment Style, Self-Protective Beliefs, and Feelings of Rejection, after Controlling for Gender and Cognitive Reflective Ability

Moderator:	Model 1		Model 2		Model 3		Model 4		Model 5	
	None		Preoccupied Attachment		Fearful Attachment		Self-Protective Beliefs		Feelings of Rejection	
	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>
Intercept	49.14***	2.02	48.73***	2.08	49.11***	2.11	48.35***	2.12	49.04***	2.11
Level 2—Between Individual										
Gender (Female)	-4.30*	2.15	-4.15†	2.18	-4.25*	2.16	-4.34*	2.16	-4.26†	2.17
CRA	2.34*	1.04	2.36*	1.05	2.40*	1.03	2.33*	1.04	2.39*	1.04
BPD	0.17	1.49	0.09	1.71	0.56	1.92	-1.67	1.97	-0.43	1.75
Moderator			-1.26	1.59	-1.01	1.62	0.81	1.71	1.31	1.65
BPD * Moderator			0.98	1.46	-0.01	1.54	1.74	1.32	0.19	1.40
Level 1—Within Individual										
Time	-0.08	0.57	0.11	0.60	-0.12	0.62	0.31	0.62	0.09	0.60
DT	-20.74***	2.40	-21.41***	2.52	-20.61***	2.59	-20.85***	2.58	-21.42***	2.52
DS	-0.76	1.16	-0.64	1.21	-0.73	1.24	-1.12	1.25	-0.64	1.23
RT	10.55***	1.86	9.49***	1.93	9.57***	2.00	10.21***	2.02	10.11***	1.97
RS	1.29**	0.40	1.32**	0.42	1.34**	0.43	1.16**	0.43	1.32**	0.42
Time * BPD	-1.12†	0.57	-1.20†	0.65	-1.14	0.74	0.08	0.75	-0.32	0.66
Time * Moderator			1.29*	0.61	-0.04	0.62	-1.08†	0.65	-1.49*	0.62
Time * BPD * Moderator			-0.63	0.56	0.07	0.59	-0.81	0.50	-0.44	0.53
DT * BPD	2.13	2.40	1.56	2.72	1.44	3.09	-1.03	3.13	-1.17	2.75
DT * Moderator	-	-	-1.93	2.54	2.16	2.62	6.30*	2.72	6.18*	2.61
DT * BPD * Moderator	-	-	2.37	2.34	-0.24	2.48	0.23	2.10	1.80	2.22
DS * BPD	2.33*	1.16	3.14*	1.31	3.56*	1.48	2.08	1.52	3.06*	1.34
DS * Moderator	-	-	-1.29	1.22	-2.53*	1.25	-0.82	1.32	-1.47	1.27
DS * BPD * Moderator	-	-	-0.76	1.13	-0.33	1.19	0.76	1.02	-0.32	1.08
RT * BPD	-3.83*	1.85	-6.38**	2.09	-6.46**	2.38	-4.25†	2.45	-4.98*	2.15
RT * Moderator	-	-	3.31†	1.95	2.89	2.02	-0.38	2.13	1.40	2.05
RT * BPD * Moderator	-	-	3.40†	1.79	2.45	1.91	0.72	1.64	1.15	1.74
RS * BPD	0.43	0.40	0.48	0.45	0.44	0.51	0.05	0.52	0.55	0.46
RS * Moderator	-	-	-0.10	0.42	0.19	0.43	0.34	0.45	-0.18	0.44
RS * BPD * Moderator	-	-	-0.07	0.39	-0.13	0.41	0.26	0.35	-0.09	0.37
Variance Components										
	<i>Variance</i>	<i>SD</i>	<i>Variance</i>	<i>SD</i>	<i>Variance</i>	<i>SD</i>	<i>Variance</i>	<i>SD</i>	<i>Variance</i>	<i>SD</i>
Intercept	135.12	11.62	144.73	12.03	138.71	11.78	138.80	11.78	141.70	11.90
Time	15.68	3.96	15.41	3.92	15.90	3.99	14.63	3.82	14.07	3.75
DT	554.81	23.55	563.30	23.73	566.50	23.80	533.97	23.11	524.13	22.89
DS	39.01	6.25	38.53	6.21	35.50	5.96	36.75	6.06	36.80	6.07
RT	74.06	8.61	64.52	8.03	78.58	8.86	78.53	8.86	76.64	8.75
RS	13.75	3.71	13.96	3.74	13.91	3.73	13.81	3.72	14.01	3.74
Residual Error	561.01	23.69	561.32	23.69	563.21	23.73	561.20	23.69	560.76	23.68
Pseudo R ²	0.19		0.19		0.19		0.19		0.19	
df (no. of Individuals)	233		232		232		233		233	

†*p* < .10, **p* < .05, ***p* < .01, ****p* < .001, tests are two-tailed, *n*=234 participants, 3510 observations

BPD = Number of BPD traits reported on MSI-BPD. CRA = Cognitive Reflective Ability. DT = Dissolution transition. DS = Dissolution slope. RT = Restoration transition. RS = Restoration slope

BPD, cognitive reflective ability, and all four moderators were *z*-standardized and centered at the sample mean

controlled for. Parameter estimates for the change in trust are provided in Table 3. Model 1 includes only BPD as a predictor of trust, while the remaining models reflect trust as a function of BPD and the following variables: preoccupied attachment style (Model 2), fearful attachment style (Model 3), self-protective beliefs (Model 4), and baseline feelings of rejection (Model 5)

Formation Slope

Growth in trust during the formation stage is represented by the estimates associated with Time. As seen in Model 1 in Table 3, individuals higher in BPD are associated with a decrease in trust growth during the formation stage (Model 1: $est_{Time * BPD} = -1.12, p < .10$). This effect is

visualized in Figure 2 in an earlier publication (Abramov et al., 2020), whereby individuals with higher BPD experienced a decreasing linear trend during the formation stage. This effect was present when controlling for preoccupied attachment style (Model 2: $est_{Time \cdot BPD} = -1.20, p < .10$) and presented a trend towards significance after controlling for fearful attachment style (Model 3: $est_{Time \cdot BPD} = -1.14, p = .12$). However, BPD no longer significantly predicted the rate of trust growth during the formation stage after controlling for self-protective beliefs (Model 4: $est_{Time \cdot BPD} = 0.08, n.s.$) or baseline feelings of rejection (Model 5: $est_{Time \cdot BPD} = -0.34, n.s.$).

Preoccupied attachment style predicted a positive linear trend in MUs transferred during formation after controlling for BPD (Model 2: $est_{Time \cdot Preoccupied} = 1.29, p < .05$). However, preoccupied attachment was not found to significantly moderate the effect of BPD on the rate of MUs transferred during trust formation (Model 2: $est_{Time \cdot BPD \cdot Preoccupied} = -0.63, n.s.$). Conversely, there was no statistical influence of fearful attachment on the growth of trust (Model 3: $est_{Time \cdot Fearful} = -0.04, n.s.$) nor evidence to suggest that fearful attachment moderates the relationship between BPD and the rate of MUs transferred (Model 3: $est_{Time \cdot BPD \cdot Fearful} = 0.07, n.s.$). Self-protective beliefs were associated with a gradual decrease in the number of MUs transferred, after controlling for the influence of BPD (Model 4: $est_{Time \cdot Protection} = -1.09, p < .10$), and there was a trend towards self-protective beliefs moderating the effect of BPD on the MU transfer rate during trust formation (Model 4: $est_{Time \cdot BPD \cdot Protection} = -0.81, p = .11$). As illustrated in Fig. 1, higher levels of self-protective beliefs may exacerbate the negative effect of BPD on the rate of MUs transferred during the formation stage.

Finally, baseline feelings of rejection was significantly associated with declining trust during the formation phase

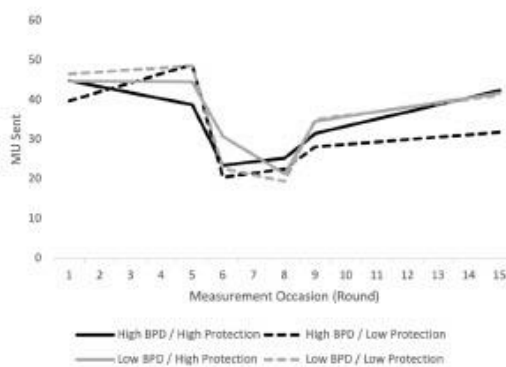


Fig. 1 Influence of BPD on Trust Growth Moderated by Self-Protective Beliefs

(Model 5: $est_{Time \cdot Rejection} = -1.49, p < .05$), but there was no evidence to suggest that baseline feelings of rejection moderated the relationship between BPD and the rate of MUs transferred (Model 5: $est_{Time \cdot BPD \cdot Rejection} = -0.44, n.s.$).

Dissolution Transition

BPD did not significantly predict the number of MUs transferred immediately following the trust violation, and this remained the case after accounting for attachment style, self-protective beliefs, and increased feelings of rejection in respective models (see Models 1-5 in Table 3). Self-protective beliefs (Model 4: $est_{DT \cdot Protection} = 6.30, p < .05$) and baseline feelings of rejection (Model 5: $est_{DT \cdot Rejection} = 6.18, p < .05$) were both found to predict a less pronounced decrease in the number of MUs transferred in response to the initial trust violation, after accounting for BPD. However, no significant three-way interactions between the dissolution transition, BPD, and each of the moderator variables were observed (see Models 2-5 in Table 3).

Dissolution Slope

As seen in Table 3, BPD was found to significantly predict a faster rate of growth in the number of MUs transferred during the dissolution phase (Model 1: $est_{DS \cdot BPD} = 2.33, p < .05$). See Figure 2 in an earlier publication for a visual representation (Abramov et al., 2020). Significance was maintained when attachment style and baseline feelings of rejection were accounted for in respective models (see Models 2, 3, and 5 in Table 3). When self-protective beliefs were accounted for, BPD no longer significantly influenced trust growth during the dissolution phase (see Model 4 in Table 3). Only fearful attachment style significantly predicted the rate of MUs transferred during the dissolution phase after accounting for number of BPD symptoms and had a negative effect on growth (Model 3: $est_{DS \cdot Fearful} = -2.53, p < .05$). No significant three-way interactions between the dissolution slope, BPD, and each of the moderator variables were observed (see Models 2-5 in Table 3).

Restoration Transition

BPD was found to significantly predict how many MUs were transferred immediately following the first instance of trust repair (Model 1: $est_{RT \cdot BPD} = -3.83, p < .05$), by reducing the size of the investment. After controlling for each of the moderator variables, this effect continued to be significant. Only preoccupied attachment style was found to predict the size of the restoration transition after accounting for BPD, but in contrast to BPD, the effect was positive (Model 2: $est_{RT \cdot Preoccupied} = 3.31, p < .10$). Preoccupied attachment

was also found to significantly moderate the effect of BPD on the rate of MUs transferred at the restoration transition (Model 2: $\text{est}_{\text{Time} \times \text{BPD} \times \text{Moderator}} = 3.40, p < .10$). As illustrated in Fig. 2, results suggest that higher levels of preoccupied attachment counteract the negative effect of BPD on the rate of MUs transferred at the restoration transition.

Restoration Slope

BPD did not significantly predict the rate of growth in the number of MUs transferred during the restoration phase, and this remained the case after accounting for attachment style, self-protective beliefs, and baseline feelings of rejection in separate models (see Models 1-5 in Table 3). Additionally, none of the moderator variables was found to predict trust growth during restoration after controlling for BPD, and no significant three-way interactions between the restoration slope, BPD, and each of the moderator variables were observed (see Models 2-5 in Table 3).

Discussion

Summary

The current study used an experimental trust game to examine whether attachment style, self-protective beliefs, and feelings of rejection explain the paradoxical relationship between BPD and trust (Abramov et al., 2020). Results suggest that the relationship between borderline pathology and the trajectory of trust growth when it is forming appears to be influenced by a number of these variables. Endorsing the belief that one needs to take action to protect oneself appears to underlie and possibly reinforce, the effect of BPD on the way trust is formed and potentially dissolved. First, when

such beliefs were controlled for, the decline in trust during the formation phase associated with BPD trait count was no longer significant. Second, there was evidence of a non-significant trend that the gradual decline in trust during the initial phase of the interaction found to be associated with borderline pathology, appeared to become more pronounced as self-protective beliefs increased. Third, self-protective beliefs were associated with a smaller decline in funds sent after the initial trust violation. Moreover, the paradoxical growth in trust in response to multiple, consecutive trust violations that was associated with borderline pathology no longer reached significance when level of self-protective beliefs was taken into account.

Given the relationship between BPD and the expectation of betrayal and abuse by others (Barazandeh et al., 2016; Bhar et al., 2008), it is possible that individuals with a high number of BPD traits entered the game already believing that the other player was untrustworthy and likely to betray. These beliefs may activate the use of self-protecting behaviours, reflected in the increasingly smaller amounts invested during the formation phase. This is also in line with previous research where, compared to healthy controls, individuals with BPD tended to be more pessimistic when predicting trust game outcomes even in the absence of any feedback regarding trustee reciprocity (Unoka et al., 2009). The study authors suggested the lowered expectancies were related to more general beliefs about the trustworthiness of others, rather than beliefs specific to their trust game partner. Indeed, there is evidence that BPD trait count is not associated with perceptions of trust game partner trustworthiness or fairness (Abramov et al., 2020), and individuals with BPD have not been found to differ from healthy controls in the accuracy of their appraisals of trustee fairness (Franzen et al., 2011).

This gives weight to the notion that the mistrustful behaviours exhibited during the beginning of the trust game despite engaging with a cooperative partner, are self-protective responses shaped by past experiences of betrayal, rather than reactions to the actual trustworthiness (reciprocity) of the partner or systematic differences in how reciprocity levels are appraised as indicators of trustworthiness. As such, holding self-protective beliefs may override the experience of an objectively cooperative partner, and it is even possible that the partner's repetitive cooperative exchanges make the anticipated betrayal more salient. Waiting for the 'inevitable' breach of trust with a new, cooperative partner, much like waiting for the other shoe to drop, may then be associated with taking self-protective action to first mitigate the risk by investing defensively, before then trying to salvage the relationship with a less negatively reactive response to the initial trust violation. That is, perhaps the mistrustful stance observed during the formation phase was

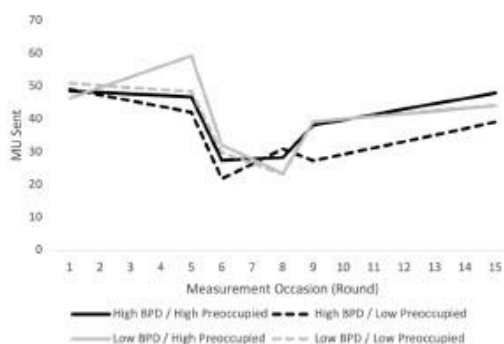


Fig. 2 Influence of BPD on Trust Growth Moderated by Preoccupied Attachment Style

a conscious strategy to obtain evidence that the other will betray, and the initial betrayal having validated the expected 'rules of engagement', allowed the high-BPD trait individual to attempt to reengage the trustee by making progressively larger investments. In line with this, De Panfilis et al., (2019) found that in other economic games, BPD patients were more likely than controls to punish their partners when receiving fair offers, but this was not the case when they received unfair offers.

As observed with self-protective beliefs, existing feelings of rejection on trust growth during the initial phase of the game appears to exert a suppressant effect on the rate of trust formation, which potentially explains the declining trust associated with high levels of BPD traits. While rejection sensitivity was not explicitly measured, these results were consistent with previous findings that sensitivity to rejection may mediate the relationship between BPD and negatively biased appraisals of trustworthiness (Miano et al., 2013; Richetin et al., 2018). Greater feelings of rejection were also associated with a less pronounced fall in trust in response to the initial trust violation which, like self-protective beliefs, may reflect an expectation of and therefore less reactivity to rejection or betrayal by the other.

While self-protective beliefs and feelings of rejection appear to reinforce some of the trust patterns observed in individuals with high levels of BPD traits, the findings on attachment insecurity suggest that the influence of borderline pathology on trust patterns occurs *despite* the contrasting effects of attachment insecurity. For example, while borderline pathology was associated with greater mistrust when trust was forming and in response to the initial trust repair effort, a preoccupied attachment style was associated with a faster rate of trust formation, and a more generous response to initial repair. This is in line with previous empirical findings that attachment anxiety is associated with increased interpersonal anxiety in response to affiliative overtures from a potential close other, a strong preference to making affiliative overtures, and a preoccupation with reciprocity (Bartz & Lydon, 2006, 2008; Fett et al., 2016). In fact, as preoccupied attachment increased, the negative effect of BPD on trust behaviours in response to the initial trust repair became less pronounced, suggesting preoccupied attachment style may reflect a more socially normative style of relating, whereby cooperation is rewarded and uncooperativeness is punished, as observed in general populations (Fulmer & Gelfand, 2015; Korsgaard et al., 2018).

In contrast, whereas borderline pathology was associated with a paradoxical growth of trust in response to multiple, consecutive trust violations, individuals reporting higher levels of fearful attachment responded by gradually reducing trust. This may have reflected the importance of reciprocity among those reporting greater levels of attachment

anxiety, although curiously, it was fearful rather than preoccupied attachment that reached significance. However overall, these findings suggest that attachment insecurity does not appear to explain the paradoxical increase in trust behaviours in high BPD trait individuals. These findings suggest that despite the theoretical and empirical links between BPD and preoccupied and fearful attachment styles, attachment insecurity was not found to drive the unusual trust patterns observed in the high BPD trait individuals and may in fact have tempered these patterns.

Limitations & Future Directions

The current study has several limitations. First, a clinical sample of individuals with BPD was not used. Although relational disturbances in nonclinical samples of people with borderline personality symptoms are almost as acute as in clinical samples (Tolpin et al., 2004), it cannot be assumed that individuals with a clinical diagnosis would respond similarly. It is notable however, that despite the modest proportion of participants endorsing a clinically relevant number of BPD traits, there was nevertheless sufficient evidence to suggest these variables play a role in the relationship between BPD and trust growth.

Second, the use of anonymous one-time encounters may not have been sufficient to evoke attachment-salient responses. Previous researchers examining trust have attempted to activate the underlying attachment system by using romantic partners and engaging them in relationship threatening dialogue (Miano et al., 2017) or through the administration of oxytocin (Bartz et al., 2011). Although a trust violation was created with the view that it might elicit an attachment relevant behavioural response, it cannot be assumed that it was sufficient. However, by using unknown partners, the current study offers an important insight into how borderline pathology might influence how new relationships are developed. Finally, a simulated trustee was used rather than an actual human being. While this methodology allowed for the systematization of trustee reciprocity, there is prior evidence that the effect of attachment on cooperation in social contexts was only observed when the participants were partnered with a human rather than a computer (Taheri et al., 2018). While efforts were made to create the illusion that participants were playing against a human in the study, the possibility remains that the deception was not effective. It is possible that effects may be artificially suppressed.

This research augmented previous trust game studies by examining how attachment style, self-protective beliefs, and feelings of rejection accounted for or modified the relationship between BPD and trust behaviours using an economic exchange game. The cross-disciplinary methodology

provides a prototype for future research looking at interpersonal disturbances as dynamic, dyadic processes. The sample provided a conservative estimate of the complex interaction between BPD symptoms and multiple sources of interpersonal disturbances on trust formation, dissolution, and restoration in a discrete social exchange. It is important for future research to replicate these findings in a clinical sample, as well as examine trust processes for individuals with BPD within other dyads such as friends, family, and romantic unions, to determine whether the findings apply in these more specific and 'higher stakes' relationships.

In this study trust behaviours were quantified to explore how various features of the borderline personality might influence the relationship between BPD and trust, to make inferences about what social cognition processes might be altered in individuals with a borderline presentation. The most compelling finding was that the presence of self-protective beliefs and feelings of rejection may have a detrimental effect on how individuals with BPD traits trust in interpersonal exchanges. A potential next step would be to augment these findings by using qualitative methodology to examine how individuals make sense of their own and their partner's behaviour in such an interaction (Sharp et al., 2011). This process, known as mentalization (Allen et al., 2008; Bateman & Fonagy, 2004), has been found to be compromised among people with BPD (for reviews see Dinsdale & Crespi, 2013; Jeung & Herpertz, 2014; Lazarus et al., 2014; Mitchell et al., 2014; Richman & Unoka, 2015). Meta-analytical findings suggest the deficits observed in BPD are not decoding impairments, but rather relate to the process of reasoning (Németh et al., 2018), that is, reasoning about others' mental states in order to explain or predict behaviour (Sabbagh, 2004). Examining reasoning would allow researchers to better understand how individuals with BPD experience the mutability of interpersonal exchanges and explore in greater depth how feelings of rejection and beliefs about the need to protect against betrayal by others may compromise the development of trust.

Conclusions

This study built on previous research using discontinuous growth modelling with a trust game to examine the how attachment style, self-protective beliefs, and feelings of rejection might underpin or modify the anomalous trust patterns associated with BPD. The findings reveal that the slower rate that trust is formed among high BPD trait individuals may be underpinned or accompanied by feelings of rejection and beliefs that others will betray so one should act pre-emptively to protect oneself. Attachment insecurity appears to have an influence on trust that is in direct contrast

to the effect that BPD exerts. Finally, BPD remained a robust predictor of faster trust growth in response to the trust violation, potentially being explained by self-protective beliefs, suggesting that there is something unique to the borderline experience that creates paradoxical ways of responding in trust-based situations. Regrettably, this manner of interacting is likely to elicit and compound the interpersonal difficulties such individuals face. The findings highlight the complex nature of borderline relational disturbance, and the need for research that can assess and quantify the internal experience of individuals with BPD to explain the aberrant style of social exchange associated with this personality disorder.

Acknowledgements The authors would like to thank Professors Paul Bliese and Jonas Lang for statistical consultation, and Yossef Abramov for developing the online trust game platform.

Authors' contribution Author 1 developed the study concept and study design and performed data collection. Author's 1 and 2 performed the data analysis and interpretation. Author 1 drafted the paper, and author's 2, 3, and 4 provided critical revisions. All authors approved the final version of the paper for submission.

Funding This work was supported by postgraduate student funding from the University of Wollongong.

Open Access funding enabled and organized by CAUL and its Member Institutions

Declarations & Statements

Conflict of Interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethics Approval The study was approved by the University of Wollongong ethics committee (HE2017/253).

Consent Informed consent was obtained from all individual participants included in the study.

Data, Materials &/or Code Availability The dataset used in this study is held within the Open Science Framework repository and can be accessed at https://osf.io/e35q4/?view_only=746d0f3254444ead80081697410395cb. Dataset has not been cited in the current manuscript's reference list due to the double-blind peer review, but on acceptance the DOI and all relevant details will be added for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright

holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abramov, G., Miellet, S., Kautz, J., Grenyer, B. F. S., & Deane, F. P. (2020). The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair. *PLoS One*, 15(7), e0236170. doi:<https://doi.org/10.1371/journal.pone.0236170>
- Agrawal, H. R., Gunderson, J. G., Holmes, B. M., & Lyons-Ruth, K. (2004). Attachment Studies with Borderline Patients: A Review. *Harvard Review of Psychiatry*, 12(2), 94–104. doi:<https://doi.org/10.1080/10673220490447218>
- Allen, J., Fonagy, P., & Bateman, A. (2008). *Mentalizing in clinical practice*
- Almakias, S., & Weiss, A. (2012). Ultimatum Game behavior in light of attachment theory. *Journal of Economic Psychology*, 33(3), 515–526. doi:<https://doi.org/10.1016/j.joep.2011.12.012>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Washington, DC: Author
- Barazandeh, H., Kissane, D. W., Saeedi, N., & Gordon, M. (2016). A systematic review of the relationship between early maladaptive schemas and borderline personality disorder/traits. *Personality and Individual Differences*, 94, 130–139. doi:<https://doi.org/10.1016/j.paid.2016.01.021>
- Bartholomew, K., & Horowitz, L. M. (1991). Attachment Styles Among Young Adults: A Test of a Four-Category Model. *Journal of Personality and Social Psychology*, 61(2), 226–244
- Bartz, J. A., & Lydon, J. E. (2006). Navigating the interdependence dilemma: Attachment goals and the use of communal norms with potential close others. *Journal of Personality and Social Psychology*, 91(1), 77–96. doi:<https://doi.org/10.1037/0022-3514.91.1.77>
- Bartz, J. A., & Lydon, J. E. (2008). Relationship-specific attachment, risk regulation, and communal norm adherence in close relationships. *Journal of Experimental Social Psychology*, 44(3), 655–663. doi:<https://doi.org/10.1016/j.jesp.2007.04.003>
- Bartz, J. A., Simeon, D., Hamilton, H., Kim, S., Crystal, S., Braun, A., & Hollander, E. (2011). Oxytocin can hinder trust and cooperation in borderline personality disorder. *Social Cognitive and Affective Neuroscience*, 6(5), 556–563. doi:<https://doi.org/10.1093/scan/nsq085>
- Bateman, A., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: Mentalization based treatment*. Oxford: Oxford University Press
- Beck, A. T., & Beck, J. S. (1991). *The Personality Belief Questionnaire*. Unpublished manuscript: University of Pennsylvania
- Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, 31(1), 64–79. doi:<https://doi.org/10.1016/j.joep.2009.10.001>
- Berenson, K. R., Gregory, W. E., Glaser, E., Romirowsky, A., Rafaleli, E., Yang, X., & Downey, G. (2016). Impulsivity, Rejection Sensitivity, and Reactions to Stressors in Borderline Personality Disorder. *Cognitive Therapy and Research*, 40(4), 510–521. doi:<https://doi.org/10.1007/s10608-015-9752-y>
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1), 122. doi:<https://doi.org/10.1006/game.1995.1027>
- Bhar, S. S., Brown, G. K., & Beck, A. T. (2008). Dysfunctional beliefs and psychopathology in borderline personality disorder. *Journal of Personality Disorders*, 22(2), 165–177. doi:<https://doi.org/10.1521/pedi.2008.22.2.165>
- Bliese, P. D., Kautz, J., & Lang, J. W. (2020). Discontinuous growth models: illustrations, recommendations, and an R function for generating the design matrix. In Y. Griep & S. D. Hansen (Eds.), *Handbook on the Temporal Dynamics of Organizational Behavior* (pp. 319–350). Edward Elgar Publishing. doi:<https://doi.org/10.4337/9781788974387.00030>
- Bliese, P. D., & Lang, J. W. B. (2016). Understanding relative and absolute change in discontinuous growth models: Coding alternatives and implications for hypothesis testing. *Organizational Research Methods*, 19(4), 562–592. doi:<https://doi.org/10.1177/1094428116633502>
- Bouchard, S., Sabourin, S., Lussier, Y., & Villeneuve, E. (2009). Relationship quality and stability in couples when one partner suffers from borderline personality disorder. *Journal of Marital and Family Therapy*, 35(4), 446–455. doi:<https://doi.org/10.1111/j.1752-0606.2009.00151.x>
- Bowlby, J. (1969). *Attachment and loss / by John Bowlby*. Hogarth Press
- Bowlby, J. (1973). *Attachment and loss. Vol. II: Separation: anxiety and anger*. New York: Basic Books
- Bowlby, J. (1980). *Attachment and loss. Vol. III: Sadness and depression*. New York: Basic Books
- Butler, A. C., Brown, G. K., Beck, A. T., & Grisham, J. R. (2002). Assessment of dysfunctional beliefs in borderline personality disorder. *Behaviour Research and Therapy*, 40(10), 1231–1240. doi:[https://doi.org/10.1016/S0005-7967\(02\)00031-1](https://doi.org/10.1016/S0005-7967(02)00031-1)
- Chapman, A. L., Dixon-Gordon, K. L., Butler, S. M., & Walters, K. N. (2015). Emotional reactivity to social rejection versus a frustration induction among persons with borderline personality features. *Personality Disorders: Theory, Research, and Treatment*, 6(1), 88–96. doi:<https://doi.org/10.1037/per0000101>
- Clifton, A., Pilkonis, P. A., & McCarty, C. (2007). Social networks in borderline personality disorder. *Journal of Personality Disorders*, 21(4), 434–441. doi:<https://doi.org/10.1521/pedi.2007.21.4.434>
- Corngnet, B., Espin, A. M., Hernán-González, R., Kujal, P., & Rassenti, S. (2016). To trust, or not to trust: Cognitive reflection in trust games. *Journal of Behavioral and Experimental Economics*, 64, 20–27. doi:<https://doi.org/10.1016/j.socsc.2015.09.008>
- De Dreu, C. K. W. (2012). Oxytocin modulates the link between adult attachment and cooperation through reduced betrayal aversion. *Psychoneuroendocrinology*, 37(7), 871–880. doi:<https://doi.org/10.1016/j.psyneuen.2011.10.003>
- De Panfilis, C., Schito, G., Generali, I., Ossola, P., Marchesi, C., Greccucci, A., & Gozzi, L. A. (2019). Emotions at the border: Increased punishment behavior during fair interpersonal exchanges in borderline personality disorder. *Journal of Abnormal Psychology*, 128(2), 162–172. doi:<https://doi.org/10.1037/abn0000404>
- Dinsdale, N., & Crespi, B. J. (2013). The borderline empathy paradox: Evidence and conceptual models for empathic enhancements in borderline personality disorder. *Journal of Personality Disorders*, 27(2), 172–195
- Downey, G., & Feldman, S. I. (1996). Implications of Rejection Sensitivity for Intimate Relationships. *Journal of Personality and Social Psychology*, 70(6), 1327–1343. doi:<https://doi.org/10.1037/0022-3514.70.6.1327>
- Dziobek, I., Preifler, S., Grozdanovic, Z., Heuser, I., Heekeren, H. R., & Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *Neuroimage*, 57(2), 539–548. doi:<https://doi.org/10.1016/j.neuroimage.2011.05.005>
- Ebert, A., Kolb, M., Heller, J., Edel, M. A., Roser, P., & Brüne, M. (2013). Modulation of interpersonal trust in borderline personality disorder by intranasal oxytocin and childhood trauma. *Social*

- Neuroscience*, 8(4), 305–313. doi:<https://doi.org/10.1080/17470919.2013.807301>
- Fett, A. K. J., Shergill, S. S., Korver-Nieberg, N., Yakub, F., Gromann, P. M., & Krabbendam, L. (2016). Learning to trust: Trust and attachment in early psychosis. *Psychological Medicine*, 46(7), 1437–1447. doi:<https://doi.org/10.1017/S0033291716000015>
- Fonagy, P., Steele, M., Steele, H., Leigh, T., Kennedy, R., Mattoon, G., & Gerber, A. (1995). Attachment, the reflective self, and borderline states: the predictive specificity of the Adult Attachment Interview and pathological emotional development. In R. S. Goldberg (Ed.), *Attachment theory: social, developmental, and clinical perspectives* (pp. 233–278). New York: Analytic
- Foxhall, M., Hamilton-Giachritsis, C., & Button, K. (2019). The link between rejection sensitivity and borderline personality disorder: A systematic review and meta-analysis. *British Journal of Clinical Psychology*, 58(3), 289–326. doi:<https://doi.org/10.1111/bjc.12216>
- Fraleigh, R. C., Hudson, N. W., Heffernan, M. E., & Segal, N. (2015). Are adult attachment styles categorical or dimensional? A taxometric analysis of general and relationship-specific attachment orientations. *Journal of Personality and Social Psychology*, 109(2), 354–368. doi:<https://doi.org/10.1037/pspp0000027>
- Franzen, N., Hagenhoff, M., Baer, N., Schmidt, A., Mier, D., Sammer, G., & Lis, S. (2011). Superior ‘theory of mind’ in borderline personality disorder: An analysis of interaction behavior in a virtual trust game. *Psychiatry Research*, 187(1–2), 224–233. doi:<https://doi.org/10.1016/j.psychres.2010.11.012>
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19(4), 25. doi:<https://doi.org/10.1257/089533005775196732>
- Fulmer, C. A., & Gelfand, M. J. (2013). How do I trust thee? Dynamic trust patterns and their individual and social contextual determinants. In K. Sycara, M. Gelfand, & A. Abbe (Eds.), *Models for intercultural collaboration and negotiation* (pp. 97–131). Dordrecht: Springer Netherlands
- Fulmer, C. A., & Gelfand, M. J. (2015). Trust after violations: Are collectivists more or less forgiving? *Journal of Trust Research*, 5(2), 109. doi:<https://doi.org/10.1080/21515581.2015.1051050>
- Gadassi, R., Snir, A., Berenson, K., Downey, G., & Rafaeli, E. (2014). Out of the frying pan, into the fire: Mixed affective reactions to social proximity in borderline and avoidant personality disorders in daily life. *Journal of Abnormal Psychology*, 123(3), 613–622. doi:<https://doi.org/10.1037/a0036929>
- Gunderson, J. G., & Lyons-Ruth, K. (2008). BPD’s interpersonal hypersensitivity phenotype: A gene-environment-developmental model. *Journal of Personality Disorders*, 22(1), 22–41. doi:<https://doi.org/10.1521/pedi.2008.22.1.22>
- Gunderson, J. G., Stout, R. L., McGlashan, T. H., Shea, M. T., Morey, L. C., Grilo, C. M., & Skodol, A. E. (2011). Ten-year course of borderline personality disorder: psychopathology and function from the Collaborative Longitudinal Personality Disorders study. *Archives of General Psychiatry*, 68(8), 827–837. doi:<https://doi.org/10.1001/archgenpsychiatry.2011.37>
- Jeung, H., & Herpertz, S. C. (2014). Impairments of interpersonal functioning: Empathy and intimacy in borderline personality disorder. *Psychopathology*, 47(4), 220–234. doi:<https://doi.org/10.1159/000357191>
- Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, 32(5), 865–889. doi:<https://doi.org/10.1016/j.joep.2011.05.007>
- King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, R. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science*, 321(5890), 806–810. doi:<https://doi.org/10.1126/science.1156902>
- Korsgaard, A. M., Kautz, J., Bliese, P. D., Samson, K., & Kostyszyn, P. (2018). Conceptualising time as a level of analysis: New directions in the analysis of trust dynamics. *Journal of Trust Research*, 8(2), 142–165. doi:<https://doi.org/10.1080/21515581.2018.1516557>
- Lazarus, S. A., Beeney, J. E., Howard, K. P., Strunk, D. R., Pilkonis, P. A., & Cheavens, J. S. (2020). Characterization of relationship instability in women with borderline personality disorder: A social network analysis. *Personality Disorders: Theory, Research, and Treatment*, 11(5), 312–320. doi:<https://doi.org/10.1037/per0000380>
- Lazarus, S. A., Cheavens, J. S., Festa, F., & Rosenthal, M. Z. (2014). Interpersonal functioning in borderline personality disorder: A systematic review of behavioral and laboratory-based assessments. *Clinical Psychology Review*, 34(3), 193–205. doi:<https://doi.org/10.1016/j.cpr.2014.01.007>
- Lazarus, S. A., Cheavens, J. S., Festa, F., & Zachary Rosenthal, M. (2014). Interpersonal functioning in borderline personality disorder: A systematic review of behavioral and laboratory-based assessments. *Clinical Psychology Review*, 34(3), 193–205. doi:<https://doi.org/10.1016/j.cpr.2014.01.007>
- Liebke, L., Koppe, G., Bungert, M., Thome, J., Hauschild, S., Defiebre, N., & Lis, S. (2018). Difficulties with being socially accepted: An experimental study in borderline personality disorder. *Journal of Abnormal Psychology*. doi:<https://doi.org/10.1037/abn0000373> 1037/abn0000373.supp (Supplemental)
- Linehan, M. M. (1993). *Cognitive-behavioural treatment of borderline personality disorder*. New York: Guilford Press
- McChure, G., Hawes, D. J., & Dadds, M. R. (2016). Borderline personality disorder and neuropsychological measures of executive function: A systematic review. *Personality and Mental Health*, 10(1), 43–57. doi:<https://doi.org/10.1002/pmh.1320>
- McChure, M. J., Bartz, J. A., & Lydon, J. E. (2013). Uncovering and overcoming ambivalence: The role of chronic and contextually activated attachment in two-person social dilemmas. *Journal of Personality*, 81(1), 103–117. doi:<https://doi.org/10.1111/j.1467-6494.2012.00788.x>
- Miano, A., Fertuck, E. A., Arntz, A., & Stanley, B. (2013). Rejection sensitivity is a mediator between borderline personality disorder features and facial trust appraisal. *Journal of Personality Disorders*, 27(4), 442–456. doi:<https://doi.org/10.1521/pedi.2013.27.096>
- Miano, A., Fertuck, E. A., Roepke, S., & Dziobek, I. (2017). Romantic relationship dysfunction in borderline personality disorder—a naturalistic approach to trustworthiness perception. *Personality Disorders: Theory, Research, and Treatment*, 8(3), 281–286. doi:<https://doi.org/10.1037/per0000196>
- Mikulincer, M. (1998). Attachment working models and the sense of trust: An exploration of interaction goals and affect regulation. *Journal of Personality and Social Psychology*, 74(5), 1209–1224. doi:<https://doi.org/10.1037/0022-3514.74.5.1209>
- Mitchell, A. E., Dickens, G. L., & Picchioni, M. M. (2014). Facial Emotion Processing in Borderline Personality Disorder: A Systematic Review and Meta-Analysis. *Neuropsychology Review*, 24(2), 166–184. doi:<https://doi.org/10.1007/s11065-014-9254-9>
- Németh, N., Mátrai, P., Hegyi, P., Czéh, B., Czopf, L., Hussain, A., & Simon, M. (2018). Theory of mind disturbances in borderline personality disorder: A meta-analysis. *Psychiatry Research*, 270, 143–153. doi:<https://doi.org/10.1016/j.psychres.2018.08.049>
- Paret, C., Jennen-Steinmetz, C., & Schmahl, C. (2017). Disadvantageous decision-making in borderline personality disorder: Partial support from a meta-analytic review. *Neuroscience and Biobehavioral Reviews*, 72, 301–309. doi:<https://doi.org/10.1016/j.neubiorev.2016.11.019>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2019). nlme: Linear and nonlinear mixed effects models. Retrieved from <https://svn.r-project.org/R-packages/trunk/nlme>

- Poggi, A., Richetin, J., & Preti, E. (2019). Trust and Rejection Sensitivity in Personality Disorders. *Current Psychiatry Reports*, 21(8), 69. doi:<https://doi.org/10.1007/s11920-019-1059-3>
- R Development Core Team (2018). R: A Language and Environment for Statistical Computing (Version 3.6.1). Vienna: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>
- Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in Close Relationships. *Journal of Personality and Social Psychology*, 49(1), 95–112. doi:<https://doi.org/10.1037/0022-3514.49.1.95>
- Richetin, J., Poggi, A., Ricciardelli, P., Fertuck, E. A., & Preti, E. (2018). The emotional components of rejection sensitivity as a mediator between borderline personality disorder and biased appraisal of trust in faces. *Clinical Neuropsychiatry*, 15(4), 200–205
- Richman, M. J., & Unoka, Z. (2015). Mental state decoding impairment in major depression and borderline personality disorder: meta-analysis. *British Journal of Psychiatry*, 207(6), 483–489. doi:<https://doi.org/10.1192/bjp.bp.114.152108>
- Ruocco, A. C. (2005). The neuropsychology of borderline personality disorder: A meta-analysis and review. *Psychiatry Research*, 137(3), 191–202. doi:<https://doi.org/10.1016/j.psychres.2005.07.004>
- Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition*, 55(1), 209–219. doi:<https://doi.org/10.1016/j.bandc.2003.04.002>
- Sharp, C., Ha, C., & Fonagy, P. (2011). Get them before they get you: Trust, trustworthiness, and social cognition in boys with and without externalizing behavior problems. *Development and Psychopathology*, 23(2), 647–658. doi:<https://doi.org/10.1017/S0954579410000003>
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling / Tom A. B. Snijders and Roel J. Bosker*. SAGE
- Staebler, K., Renneberg, B., Stopsack, M., Fiedler, P., Weiler, M., & Roepke, S. (2011). Facial emotional expression in reaction to social exclusion in borderline personality disorder. *Psychological Medicine*, 41(9), 1929–1938. doi:<https://doi.org/10.1017/S0033291711000080>
- Stiglmayr, C. E., Grathwol, T., Linehan, M. M., Ihorst, G., Fahrenberg, J., & Bohus, M. (2005). Aversive tension in patients with borderline personality disorder: A computer-based controlled field study. *Acta Psychiatrica Scandinavica*, 111(5), 372–379. doi:<https://doi.org/10.1111/j.1600-0447.2004.00466.x>
- Taberi, M., Rotshtein, P., & Beierholm, U. (2018). The effect of attachment and environmental manipulations on cooperative behavior in the prisoner's dilemma game. *PLoS One*, 13(11), e0205730. doi:<https://doi.org/10.1371/journal.pone.0205730>
- Tolpin, L. H., Gunthert, K. C., Cohen, L. H., & O'Neill, S. C. (2004). Borderline personality features and instability of daily negative affect and self-esteem. *Journal of Personality*, 72(1), 111–137. doi:<https://doi.org/10.1111/j.0022-3506.2004.00258.x>
- Unoka, Z., Seres, I., Áspán, N., Bódi, N., & Kéri, S. (2009). Trust game reveals restricted interpersonal transactions in patients with borderline personality disorder. *Journal of Personality Disorders*, 23(4), 399–409. doi:<https://doi.org/10.1521/pedi.2009.23.4.399>
- Wolff, S., Stiglmayr, C., Bretz, H. J., Lammers, C., & Auckenthaler, A. (2007). Emotion identification and tension in female patients with borderline personality disorder. *British Journal of Clinical Psychology*, 46(3), 347–360
- Zanarini, M. C., Frankenburg, F. R., Reich, D. B., & Fitzmaurice, G. (2010). Time to attainment of recovery from borderline personality disorder and stability of recovery: A 10-year prospective follow-up study. *American Journal of Psychiatry*, 167(6), 663–667. doi:<https://doi.org/10.1176/appi.ajp.2009.09081130>
- Zanarini, M. C., Vujanovic, A. A., Parachini, E. A., Boulanger, J. L., Frankenburg, F. R., & Hennen, J. (2003). A screening measure for BPD: The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). *Journal of Personality Disorders*, 17, 568–573. doi:<https://doi.org/10.1521/pedi.17.6.568.25355>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix C: Research Materials (Chapters 2-4)

Ethics Approval	205
Brief Study Description on SONA	207
Participant Information Statement	208
Online Information and Consent Form	210
Pre-Game Questions	213
Trust Game Instructions and Test Questions	221
Trust Game Screenshots: Round 1 Example	223
Post-Game Social-Cognitive Reasoning Questions Instructions	227
Post-Game Trustee Appraisal	232
Participant Debrief	233

Ethics Approval

Dear Professor Grenyer,

I am pleased to advise that the application detailed below has been **approved**.

Ethics Number: 2017/253

Approval Date: 11/07/2017

Expiry Date: 10/07/2018

Project Title: The Investment Game Study

Researcher/s: Abramov Gamze; Bourke Marianne; Miellest Sebastien; Grenyer Brin

Documents Approved: Revised Protocol - 03/07/2017
Response to review - 03/07/2017

Sites:

Site	Principal Investigator for Site
University of Wollongong	Professor Brin Grenyer

The HREC has reviewed the research proposal for compliance with the *National Statement on Ethical Conduct in Human Research* and approval of this project is conditional upon your continuing compliance with this document. Compliance is monitored through progress reports; the HREC may also undertake physical monitoring of research.

Approval is granted for a twelve month period; extension of this approval will be considered on receipt of a progress report **prior to the expiry date**. Extension of approval requires:

- The submission of an annual progress report and a final report on completion of your project.
- Approval by the HREC of any proposed changes to the protocol or investigators.
- Immediate report of serious or unexpected adverse effects on participants.
- Immediate report of unforeseen events that might affect the continued acceptability of the project.

If you have any queries regarding the HREC review process or your ongoing approval please contact the Ethics Unit on 4221 3386 or email rso-ethics@uow.edu.au.

Yours sincerely,

Melanie Randle

Associate Professor Melanie Randle,

Chair, UOW & ISLHD Social Sciences Human Research Ethics Committee

The University of Wollongong and Illawarra and Shoalhaven Local Health District Social Sciences HREC is constituted and functions in accordance with the NHMRC National Statement on Ethical Conduct in Human Research.

Brief Study Description on SONA

It is common for mental health to impact our interpersonal relationships and the way we make decisions in interpersonal situations. There is research to suggest that our decision making can be affected by our beliefs about ourselves and others, our relationship style, our capacity to think about what someone else might be thinking or feeling, and our own feelings during interactions. We are interested in examining how these factors impact the way people make decisions involving money in an investment game.



Participant Information Statement

The Investment Game Study

1) What is the study about?

You are invited to participate in a study investigating how people invest money and the decisions they make in interpersonal situations.

2) Who is carrying out the study?

The study is being conducted by Professor Brin Grenyer, Dr Sebastien Miellet, Dr Marianne Bourke and Gamze Abramov (PhD candidate) based in the School of Psychology, University of Wollongong.

3) What does the study involve?

The study involves participating in an online game in which you choose how much money to send to the other player for investment. The study is being conducted at multiple locations in Australia and overseas. You will be paired with another participant from one of these sites anonymously. The money used is virtual money and it will not cost you anything to play. You will also be asked to complete a survey that asks you about your beliefs, mood, relationships, and personality, as well as general information about yourself.

4) Where will I participate in the study?

The study is conducted entirely online. There will be timeslots available 7 days a week, 24 hours a day during the study period. When you sign up you will be asked to book in a time to complete the study using the Sona System. You will need a PC or laptop and an internet connection.

5) How much time will the study take?

It is anticipated that participation will take approximately 1 hour.

6) How many credit points will I receive for participating in this study?

You will receive 1 credit point for participation.

7) Can I withdraw from the study?

Participation in the study is completely voluntary - you are not under any obligation to consent and - if you do consent - you can withdraw at any time without affecting your relationship with the researchers or the University of Wollongong.

8) What if I feel distressed during or after the interview?

The questions in the survey and the Investment Game have been used in multiple studies with no adverse effects reported. Nevertheless, you may experience some discomfort. Should you wish to speak to one of the researchers about your experience of the study, please contact Gamze Abramov on ga385@uowmail.edu.au to organize a time. If you do become distressed during or after participating in the study and wish to seek support, please call LifeLine on 13 11 14 (within Australia). If outside Australia, please contact your nearest doctor or hospital health service.

9) Will anyone else know the results?

All aspects of the study, including results, will be strictly confidential. A report of the study may be submitted for publication and may be presented at conferences, but individual participants will not be identifiable in such publication or presentation.

10) What if I require further information about the study or my involvement in it?

Enquiries about the research should be directed to Professor Brin Grenyer, (02) 4221 3474 (grenyer@uow.edu.au) or Gamze Abramov (ga385@uowmail.edu.au).

11) What if I have a complaint or any concerns?

This study has been reviewed by the Social Sciences Human Research Ethics Committee at the University of Wollongong. If you have any concerns or complaints regarding the way this research has been conducted, you can contact the UOW Ethics Officer on (02) 4221 3386 or email rso-ethics@uow.edu.au.

Online Information and Consent Form

PROJECT TITLE: The investment game study

RESEARCHER: Gamze Abramov, Ba (Hons), PhD (Clinical Psychology) Candidate, School of Psychology and Illawarra Health and Medical Research Institute, University of Wollongong, Australia; Email: ga385@uowmail.edu.au

SUPERVISORS: Professor Brin Grenyer, School of Psychology, Illawarra Health and Medical Research Institute, University of Wollongong, Australia; Phone: +41 2 4221 3474, Email: grenyer@uow.edu.au

Dr Sebastien Miellet, School of Psychology and Illawarra Health and Medical Research Institute, University of Wollongong, Australia; Email: smiellet@uow.edu.au

Dr Marianne Bourke, School of Psychology and Illawarra Health and Medical Research Institute, University of Wollongong, Australia; Email: mbourke@uow.edu.au

PURPOSE OF THE RESEARCH:

This research project aims to understand how people invest money and the decisions they make in interpersonal situations. This study is being conducted by Gamze Abramov as part of the Doctor of Philosophy (Clinical Psychology) program, under the supervision of Professor Brin Grenyer of the School of Psychology, University of Wollongong.

WHAT DOES THE STUDY INVOLVE?

The study involves two components: survey questions and an economic game followed by additional questions. I can complete these in one sitting which will take approximately one hour. If I elect to complete the study in two or more sittings, I understand I will need to provide an email address to receive a link to return to the study. The first component (survey questions) may be completed in more than one sitting and will take approximately 30 minutes. This survey will ask questions about my feelings and attitudes to relationships (e.g., “I worry a lot about my relationships”).

The second component (game and pre- and post-game questions) *must* be completed in *one* sitting and will take approximately 30 minutes. Both participants will remain anonymous, and the game will involve deciding how much money to give to the other player. Before playing the game and after playing the game I will be asked to complete questions about my mood and my impression of the other player.

POSSIBLE RISKS AND BENEFITS:

I understand that the questionnaires and game have been used in multiple studies with no adverse outcomes reported. I understand that some of the questions ask about relationships and mood and may cause some discomfort. I have been advised that should I want further information I am able to discuss this with the research, Gamze Abramov. I understand that participation in this research is voluntary, includes no immediate benefit to myself, however, may provide the opportunity to help us better understand the factors that impact how people make decisions in interpersonal situations. I understand that I may refuse to participate or withdraw my data at any time without consequence. If I wish to withdraw from the study whilst completing the online survey or playing the game, I simply have to exit the survey by clicking on the exit tab which will appear on every page.

I understand that all information provided will be kept strictly confidential, and no identifying information will be stored with the data. The data collected will be used for the purpose of journal publications, conference presentations, a doctoral research thesis and to help us understand what factors influence the way people make decisions in a financial setting. Summary information only will be included, and no individual will be identifiable in the reporting of results.

CONTACTS:

If I have any questions about this research, I can contact Gamze Abramov or Professor Brin Grenyer. I understand that answering personal questions can sometimes be distressing and that residents in Australia can contact LifeLine on tel: 13 11 14 for a free counselling service 24 hours per day, 7 days per week if participating in the study is distressing. If outside Australia, please contact your local doctor or hospital health service.

ETHICS REVIEW:

I understand that the ethical aspects of the study have been approved by the Social Sciences Human Research Ethics Committee at the University of Wollongong. If I have concerns or complaints regarding the way this research has been conducted, I can contact the University of Wollongong Ethics Officer on +61 2 4112 3386 or email rso-ethics@uow.edu.au

CONSENT:

If you decide to participate in this study, please complete the following questions. By clicking 'I ACCEPT' below and answering these questions you are indicating that you agree to participate in this study. If you do not wish to participate, simply close this link.

By clicking one of the buttons below I am indicating my agreement, or otherwise, to participate in this research

I accept

I do not wish to proceed

Once you have completed the survey and game you will be asked if you wish to be involved in a future studies. If you agree, you will be asked to provide an email address.

Pre-Game Questions

The first part of the study involves answering a number of different questions. Please read the instructions at the top of each page before commencing. You must answer every question before you will be allowed to progress to the next page. If you wish to withdraw from the study, you can do so by clicking the Exit tab at the top of the screen.

Demographic questions

What year were you born? {years in drop down menu}

What is your gender? {Female, Male}

McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD; Zanarini et al., 2003)

Below you will find a number of descriptive statements. Please consider each one and indicate how well that statement describes you. Please answer honestly, as there are no right or wrong answers. Simply indicate how well each statement describes your experiences.

6-point scale: (0) None of the time... (5) All of the time

1. Have any of your closest relationships been troubled by a lot of arguments or repeated breakups?
2. Have you deliberately hurt yourself physically (e.g., punched yourself, cut yourself, burned yourself)? How about made a suicide attempt?
3. Have you had at least two other problems with impulsivity (e.g., eating binges and spending sprees, drinking too much and verbal outbursts)?
4. Have you been extremely moody?
5. Have you felt very angry a lot of the time? How about often acted in an angry or sarcastic manner?
6. Have you often been distrustful of other people?
7. Have you frequently felt unreal or as if things around you were unreal?
8. Have you chronically felt empty?
9. Have you often felt that you had no idea of who you are or that you have no identity?
10. Have you made desperate efforts to avoid feeling abandoned or being abandoned (e.g., repeatedly called someone to reassure yourself that he or she still cared, begged them not to leave you, clung to them physically)?

Mental Health Inventory (MHI-5; Berwick et al., 1991)

These questions are about how you feel and how things have been with you over the past two weeks. For each question, please indicate the answer that comes closest to the way you have been feeling.

6-point scale: 0 (None of the time) ... 5 (All of the time)

1. Have you been a very nervous person?
2. Have you felt so down in the dumps that nothing could cheer you up?
3. Have you felt calm and peaceful?
4. Have you felt downhearted and blue?
5. Have you been a happy person?

*Standardized Assessment of Personality – Abbreviated Scale Self-Report (SAPAS-SR;
Germans et al., 2008)*

Indicate whether you think that the description applies most of the time and in most situations.

Yes/No

1. In general, do you have difficulty making and keeping friends?
2. Would you normally describe yourself as a loner?
3. In general, do you trust other people?
4. Do you normally lose your temper easily?
5. Are you normally an impulsive sort of person?
6. Are you normally a worrier?
7. In general, do you depend on others a lot?
8. In general, are you a perfectionist?

Cognitive Reflection Test (CRT; Frederick, 2005)

Please work out the following problems.

1. A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? _____ cents
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days

Relationship Questionnaire (RQ; Bartholomew & Horowitz, 1991)

These statements are about how you feel in your relationships with others. Please read the following passages very carefully, then rate each paragraph according to how closely the paragraph describes you.

11-point scale: (0%) Not at all like me... (100%) Very much like me

- A. It is easy for me to become emotionally close to others. I am comfortable depending on them and having them depend on me. I don't worry about being alone or having others not accept me.
- B. I am uncomfortable getting close to others. I want emotionally close relationships, but I find it difficult to trust others completely, or to depend on them. I worry that I will be hurt if I allow myself to become too close to others.⁵
- C. I want to be emotionally intimate with others, but I often find that others are reluctant to get as close as I would like. I am uncomfortable being without close relationships, but I worry that others don't value me as much as I value them.⁶
- D. I am very comfortable without close relationships. It is very important to me to feel independent and self-sufficient, and I prefer not to depend on others or have others depend on me.

⁵ Used to measure fearful attachment style

⁶ Used to measure preoccupied attachment style

Personality Beliefs Questionnaire—Borderline Personality Subscale (PBQ-BPD; Butler et al., 2002)

Each statement below describes a pattern of behaviour or thinking. Consider each statement carefully and then decide if the statement describes the way you act or think.

5-point scale: (0) I don't believe it at all... (4) I believe it totally

1. Unpleasant feelings will escalate and get out of control.
2. Any signs of tension in a relationship indicate that the relationship has gone bad; therefore, I should cut it off.⁷
3. I can't cope as other people can.
4. People will pay attention only if I act in extreme ways.⁸
5. People will get me if I don't get them first.⁹
6. I have to be on guard at all times.
7. I am helpless when left on my own.
8. A person whom I am close to could be disloyal or unfaithful.
9. If people get close to me, they will discover the real me and reject me.
10. I need somebody around available at all times to help me carry out what I need to do or in case something bad happens.
11. People will take advantage of me if I give them the chance.
12. I am needy and weak.
13. I cannot trust other people.
14. People often say one thing and mean something else.

⁷ Used to measure self-protective beliefs

⁸ Used to measure self-protective beliefs

⁹ Used to measure self-protective beliefs

Feelings of Rejection

Indicate how much you feel each emotion at the present moment.

5-point scale: (0) not at all... (4) extremely

At the present moment I feel rejected by other.

At the present moment I feel accepted by others.

At the present moment I feel abandoned by others.

At the present moment I feel my needs are being met by others.

Trust Game Instructions and Test Questions

Investment Game

Congratulations! You have completed the pre-game questions.

You will now play the Investment Game.

Below are the instructions for the game. Make sure you understand how to play before clicking Next. Good luck!

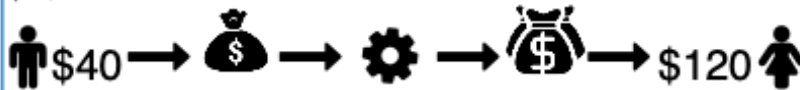
The game will take around 15 minutes and it will be followed by some more questions.

You will play this game with another player over the web. The other may be from anywhere around the world, and you will be completely anonymous to one another.

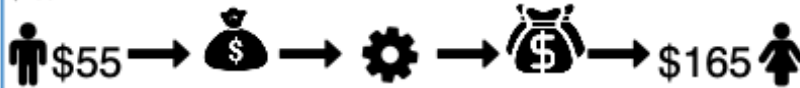
You will play multiple rounds of a game. In each round of the game you will receive \$100. You can choose to send the other player any amount between \$0 and \$100.

The system will triple any amount you send.

For example, if you send the other player \$40, the system will triple that to \$120, so the other player will receive \$120.



Similarly, if you send the other player \$55, the system will triple that to \$165, so the other player will receive \$165.



Once the other player receives money from you, they will need to make a decision of how much money, if any, to repay back to you.

Using the same examples as above, if the other player received a tripled amount of \$40, they can choose to repay any amount from \$0 to \$120.

Likewise, if the other player received a tripled amount of \$55, they can choose to repay any amount from \$0 to \$165.

Once you receive the repayment from the other player, it will be the end of that round. After each round we will tell you how much the other player repaid you.

The goal for each of you is to make as much money as possible!

You will play multiple rounds with the same partner. We will tell you when you are up to your final round.

Are you ready? If yes, click FIND ME A PARTNER.

The system will search for a partner for you. It may take up to a minute.

Good luck.

NEXT

Investment Game

Before you start we will verify that you understand the game's rules, but don't worry, if you are wrong the system will notify you and you will be able to correct to the right answer

What is the minimum amount I can send?	<input checked="" type="radio"/> \$100 \$0 \$1 <i>In each new round, you start with \$100. You can send the other player any amount from \$0 to \$100</i>
What is the maximum amount I can send?	\$50 <input checked="" type="radio"/> \$300 \$100 <i>In each new round, you start with \$100. You can send the other player any amount from \$0 to \$100</i>
If I send \$50, what will the other player receive?	\$50 <input checked="" type="radio"/> \$150 \$100
If the other player receives \$150, how much can they return to me?	<input checked="" type="radio"/> Any amount between 0 and \$150 Any amount between \$50 and \$150 Any amount between \$1 and \$150
How much money do I start off with in each new round?	Whatever I have left over from the previous round. <input checked="" type="radio"/> \$100 \$100 plus my earnings from the previous round

FIND ME A PARTNER

Please answer all the required questions in order to proceed

Trust Game Screenshots: Round 1 Example

Investment Game

Found a player, Start by sending some money

How many dollars would you like to send the other player?
0 to 100

Must be a number between 0 and 100

SEND MONEY



Investment Game

Found a player, Start by sending some money

How many dollars would you like to send the other player?

60

SEND MONEY



Investment Game

Sent: \$60

Giving the other player \$180, Waiting for reply...

Received: \$180

How many dollars would you like to send the other player?

60

SEND MONEY



Investment Game

Sent: \$60
Round 1 Balance: \$107

Received: \$180
Round 1 Balance: \$113

NEXT ROUND



Post-Game Social-Cognitive Reasoning Questions Instructions

Investment Game

Now we're going to ask you some questions about some of the choices you and the other player made during the game.

Here is a summary list of each of the investments you made, how much the other player received, and how much they repaid.

When you hit Next we will ask you about a specific round.

In round 1: you sent \$60, the other player received \$180 and repaid \$67
In round 2: you sent \$20, the other player received \$60 and repaid \$25
In round 3: you sent \$30, the other player received \$90 and repaid \$36
In round 4: you sent \$50, the other player received \$150 and repaid \$62
In round 5: you sent \$40, the other player received \$120 and repaid \$8
In round 6: you sent \$12, the other player received \$36 and repaid \$4
In round 7: you sent \$2, the other player received \$6 and repaid \$0
In round 8: you sent \$0, the other player received \$0 and repaid \$0
In round 9: you sent \$40, the other player received \$120 and repaid \$49
In round 10: you sent \$70, the other player received \$210 and repaid \$88
In round 11: you sent \$30, the other player received \$90 and repaid \$35
In round 12: you sent \$35, the other player received \$105 and repaid \$37
In round 13: you sent \$30, the other player received \$90 and repaid \$38
In round 14: you sent \$12, the other player received \$36 and repaid \$15
In round 15: you sent \$78, the other player received \$234 and repaid \$84
In round 16: you sent \$20, the other player received \$60 and repaid \$23

NEXT

Example screenshot reasoning round 1

Investment Game

Now we're going to ask you some questions about some of the choices you and the other player made during the game.
Here is a summary list of each of the investments you made, how much the other player received, and how much they repaid.
When you hit Next we will ask you about a specific round.

In round 1: you sent \$60, the other player received \$180 and repaid \$67
In round 2: you sent \$20, the other player received \$60 and repaid \$25
In round 3: you sent \$30, the other player received \$90 and repaid \$36
In round 4:
In round 5:
In round 6:
In round 7:
In round 8:
In round 9:
In round 10:
In round 11:
In round 12:
In round 13:
In round 14:
In round 15:
In round 16: you sent \$20, the other player received \$60 and repaid \$23

Round 1

In this round you sent \$60

Why did you choose to send this amount?

To see how giving they would be

NEXT

NEXT

Investment Game

Now we're going to ask you some questions about some of the choices you and the other player made during the game.

Here is a summary list of each of the investments you made, how much the other player received, and how much they repaid.

When you hit Next we will ask you about a specific round.

In round 1: you sent \$60, the other player received \$180 and repaid \$67

In round 2: you sent \$20, the other player received \$60 and repaid \$25

In round 3: you sent \$30, the other player received \$90 and repaid \$36

In round 4:

In round 5:

In round 6:

In round 7:

In round 8:

In round 9:

In round 10:

In round 11:

In round 12:

In round 13:

In round 14:

In round 15:

In round 16: you sent \$20, the other player received \$60 and repaid \$23

Round 1

In this round you sent \$60, the other player received \$180 and repaid \$67

Why do you think the other player chose to repay this amount?

To give me back more and hope I would return with more

NEXT

NEXT

Example screenshot reasoning round 3

Investment Game

Now we're going to ask you some questions about some of the choices you and the other player made during the game.
Here is a summary list of each of the investments you made, how much the other player received, and how much they repaid.
When you hit Next we will ask you about a specific round.

In round 1:	you sent \$60, the other player received \$180 and repaid \$67
In round 2:	you sent \$20, the other player received \$60 and repaid \$25
In round 3:	you sent \$30, the other player received \$90 and repaid \$30
In round 4:	you sent \$40, the other player received \$120 and repaid \$40
In round 5:	you sent \$50, the other player received \$150 and repaid \$50
In round 6:	you sent \$60, the other player received \$180 and repaid \$67
In round 7:	you sent \$70, the other player received \$210 and repaid \$70
In round 8:	you sent \$80, the other player received \$240 and repaid \$80
In round 9:	you sent \$90, the other player received \$270 and repaid \$90
In round 10:	you sent \$100, the other player received \$300 and repaid \$100
In round 11:	you sent \$110, the other player received \$330 and repaid \$110
In round 12:	you sent \$120, the other player received \$360 and repaid \$120
In round 13:	you sent \$130, the other player received \$390 and repaid \$130
In round 14:	you sent \$140, the other player received \$420 and repaid \$140
In round 15:	you sent \$150, the other player received \$450 and repaid \$150
In round 16:	you sent \$160, the other player received \$480 and repaid \$160

Round 3

In round 1 you sent \$60, the other player received \$180 and repaid \$67
In round 2 you sent \$20, the other player received \$60 and repaid \$25

In this round you sent \$30

Why did you choose to send this amount?

To test the person if he will reduce the amount

NEXT

NEXT

Investment Game

Now we're going to ask you some questions about some of the choices you and the other player made during the game.

Here is a summary list of each of the investments you made, how much the other player received, and how much they repaid.

When you hit Next we will ask you about a specific round.

In round 1: you sent \$60, the other player received \$180 and repaid \$67

In round 2: you sent \$20, the other player received \$60 and repaid \$25

In round 3: you sent \$30, the other player received \$90 and repaid \$36

In round 4:

In round 5:

In round 6:

In round 7:

In round 8:

In round 9:

In round 10:

In round 11:

In round 12:

In round 13:

In round 14:

In round 15:

In round 16:

Round 3

In round 1 you sent \$60, the other player received \$180 and repaid \$67

In round 2 you sent \$20, the other player received \$60 and repaid \$25

In this round you sent \$30, the other player received \$90 and repaid \$36

Why do you think the other player chose to repay this amount?

Because i reduced and so he did too but still wanted to show

NEXT

NEXT

Post-Game Trustee Appraisal

Rate the following on a 5-point scale (0 = "Not at all"; 4 = "Absolutely").

Did the other player play fair?

Is the other player trustworthy?

Participant Debrief

Thank you for participating in the Investment Game Study. Our study is interested in understanding how people make decisions in interpersonal situations. We used the Investment Game to examine the degree to which people cooperate with one another. There is research to suggest that this can be affected by the following factors: beliefs about others, our relationship style, our capacity to think about what someone else might be thinking or feeling, our personality traits, and our own feelings during the interaction. The questions we asked you both before and after the game measured these factors.

All the information we collected in today's study will be confidential, and there will be no way of identifying your responses in the data archive. We are not interested in any one individual's responses. Rather we want to look at the general patterns that emerge when the data are combined.

Your participation today is appreciated and will help psychologists discover more ways of promoting cooperation in interpersonal situations. We ask that you do not discuss the nature of the study with others who may later participate in it, as this could affect the validity of our research conclusions. If you have any questions or concerns, you are welcome to talk with Professor Brin Grenyer, (02) 4221 3474 (grenyer@uow.edu.au) or Gamze Abramov (ga385@uowmail.edu.au). If your participation in this study has caused you concerns, anxiety, or otherwise distressed you, you may contact Gamze Abramov on the email listed above, or call LifeLine on 13 11 14 (within Australia) or your nearest doctor or hospital health service (outside Australia).

If you are interested in learning more about the results of the study, please email your contact details to Gamze Abramov (ga385@uowmail.edu.au) to receive a copy of the findings when the study is completed.

THANK YOU AGAIN FOR YOUR PARTICIPATION.

References

- Abramov, G., Kautz, J., Miellet, S., & Deane, F. P. (2022). The influence of attachment style, self-protective beliefs, and feelings of rejection on the decline and growth of trust as a function of borderline personality disorder trait count. *Journal of Psychopathology and Behavioral Assessment*. <https://doi.org/10.1007/s10862-022-09965-9>
- Abramov, G., Miellet, S., Kautz, J., Grenyer, B. F. S., & Deane, F. P. (2020). The paradoxical decline and growth of trust as a function of borderline personality disorder trait count: Using discontinuous growth modelling to examine trust dynamics in response to violation and repair. *PLOS ONE*, *15*(7), e0236170. <https://doi.org/10.1371/journal.pone.0236170>
- Agrawal, H. R., Gunderson, J., Holmes, B. M., & Lyons-Ruth, K. (2004). Attachment studies with borderline patients: A review. *Harvard Review of Psychiatry*, *12*(2), 94–104. <https://doi.org/10.1080/10673220490447218>
- Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Erlbaum.
- Ainsworth, M. D. S., Wittig, B. A., & Foss, B. M. (1969). Attachment and exploratory behaviour of one-year-olds in a strange situation. In B. M. Foss (Ed.), *Determinants of infant behaviour* (Vol. 4, pp. 111–136). Methuen.
- Allen, J., Fonagy, P., & Bateman, A. (2008). *Mentalizing in clinical practice*. American Psychiatric Press.
- Almakias, S., & Weiss, A. (2012). Ultimatum Game behavior in light of attachment theory. *Journal of Economic Psychology*, *33*(3), 515–526. <https://doi.org/10.1016/j.joep.2011.12.012>

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.
<https://doi.org/10.1176/appi.books.9780890425596>
- Andreou, C., Kelm, L., Bierbrodt, J., Braun, V., Lipp, M., Yassari, A. H., & Moritz, S. (2015). Factors contributing to social cognition impairment in borderline personality disorder and schizophrenia. *Psychiatry Research*, *229*(3), 872–879.
<https://doi.org/10.1016/j.psychres.2015.07.057>
- Anupama, V., Bholra, P., Thirthalli, J., & Mehta, U. M. (2018). Pattern of social cognition deficits in individuals with borderline personality disorder. *Asian Journal of Psychiatry*, *33*, 105–112. <https://doi.org/10.1016/j.ajp.2018.03.010>
- Arntz, A., Weertman, A., & Salet, S. (2011). Interpretation bias in Cluster-C and borderline personality disorders. *Behaviour Research and Therapy*, *49*(8), 472–481.
<https://doi.org/10.1016/j.brat.2011.05.002>
- Ayduk, Ö., Gyurak, A., & Luerssen, A. (2008). Individual differences in the rejection–aggression link in the hot sauce paradigm: The case of rejection sensitivity. *Journal of Experimental Social Psychology*, *44*(3), 775–782.
<https://doi.org/10.1016/j.jesp.2007.07.004>
- Bach, B., & Farrell, J. M. (2018). Schemas and modes in borderline personality disorder: The mistrustful, shameful, angry, impulsive, and unhappy child. *Psychiatry Research*, *259*, 323–329. <https://doi.org/10.1016/j.psychres.2017.10.039>
- Badoud, D., Prada, P., Nicastro, R., Germond, C., Luyten, P., Perroud, N., & Debbané, M. (2018). Attachment and reflective functioning in women with borderline personality disorder. *Journal of Personality Disorders*, *32*(1), 17–30.
https://doi.org/10.1521/pedi_2017_31_283

- Baez, S., Marengo, J., Perez, A., Huepe, D., Font, F. G., Rial, V., Gonzalez-Gadea, M. L., Manes, F., & Ibanez, A. (2015). Theory of mind and its relationship with executive functions and emotion recognition in borderline personality disorder. *Journal of Neuropsychology*, *9*(2), 203–218. <https://doi.org/10.1111/jnp.12046>
- Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2013). Sniffing around oxytocin: Review and meta-analyses of trials in healthy and clinical groups with implications for pharmacotherapy. *Translational Psychiatry*, *3*, 1–14. <https://doi.org/10.1038/tp.2013.34>
- Barazandeh, H., Kissane, D. W., Saeedi, N., & Gordon, M. (2016). A systematic review of the relationship between early maladaptive schemas and borderline personality disorder/traits. *Personality and Individual Differences*, *94*, 130–139. <https://doi.org/10.1016/j.paid.2016.01.021>
- Barnow, S., Stopsack, M., Grabe, H. J., Meinke, C., Spitzer, C., Kronmüller, K., & Sieswerda, S. (2009). Interpersonal evaluation bias in borderline personality disorder. *Behaviour Research and Therapy*, *47*, 359–365. <https://doi.org/10.1016/j.brat.2009.02.003>
- Bartholomew, K., & Horowitz, L. M. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality & Social Psychology*, *61*(2), 226–244. bth.
- Barton, B. B., Goerigk, S., Wüstenberg, T., Dewald-Kaufmann, J., Reinhard, M. A., Musil, R., Ehring, T., Jobst, A., & Padberg, F. (2021). Altered immediate behavioral response to partial social exclusion: A cross-diagnostic study in patients with borderline personality disorder and persistent depressive disorder. *Journal of Psychiatric Research*, *144*, 177–183. <https://doi.org/10.1016/j.jpsychires.2021.10.005>

- Bartz, J. A., & Lydon, J. E. (2006). Navigating the interdependence dilemma: Attachment goals and the use of communal norms with potential close others. *Journal of Personality and Social Psychology, 91*(1), 77–96. <https://doi.org/10.1037/0022-3514.91.1.77>
- Bartz, J. A., & Lydon, J. E. (2008). Relationship-specific attachment, risk regulation, and communal norm adherence in close relationships. *Journal of Experimental Social Psychology, 44*(3), 655–663. <https://doi.org/10.1016/j.jesp.2007.04.003>
- Bartz, J. A., Simeon, D., Hamilton, H., Kim, S., Crystal, S., Braun, A., Vicens, V., & Hollander, E. (2011). Oxytocin can hinder trust and cooperation in borderline personality disorder. *Social Cognitive and Affective Neuroscience, 6*(5), 556–563. <https://doi.org/10.1093/scan/nsq085>
- Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: Context and person matter. *Trends in Cognitive Sciences, 15*(7), 301–309. <https://doi.org/10.1016/j.tics.2011.05.002>
- Bateman, A., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: Mentalization based treatment*. Oxford University Press. <https://doi.org/10.1093/med:psych/9780198527664.001.0001>
- Beck, A. T., & Beck, J., S. (1991). *The Personality Belief Questionnaire*. The Beck Institute for Cognitive Therapy and Research. <https://beckinstitute.org/wp-content/uploads/2021/06/PBQ-Full-Documents-1.pdf>
- Becker, D. F., McGlashan, T. H., & Grilo, C. M. (2006). Exploratory factor analysis of borderline personality disorder criteria in hospitalized adolescents. *Comprehensive Psychiatry, 47*(2), 99–105. <https://doi.org/10.1016/j.comppsy.2005.07.003>

- Beeney, J. E., Hallquist, M. N., Clifton, A. D., Lazarus, S. A., & Pilkonis, P. A. (2018). Social disadvantage and borderline personality disorder: A study of social networks. *Personality Disorders, 9*(1), 62–72. <https://doi.org/10.1037/per0000234>
- Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology, 31*(1), 64–79. <https://doi.org/10.1016/j.joep.2009.10.001>
- Berenson, K. R., Dochat, C., Martin, C. G., Yang, X., Rafaeli, E., & Downey, G. (2018). Identification of mental states and interpersonal functioning in borderline personality disorder. *Personality Disorders, 9*(2), 172–181. <https://doi.org/10.1037/per0000228>
- Berenson, K. R., Downey, G., Rafaeli, E., Coifman, K. G., & Leventhal Paquin, N. (2011). The rejection-rage contingency in borderline personality disorder. *Journal of Abnormal Psychology, 120*(3), 681–690. <https://doi.org/10.1037/a0023335>
- Berenson, K. R., Gregory, W. E., Glaser, E., Romirowsky, A., Rafaeli, E., Yang, X., & Downey, G. (2016). Impulsivity, rejection sensitivity, and reactions to stressors in borderline personality disorder. *Cognitive Therapy and Research, 40*(4), 510–521. <https://doi.org/10.1007/s10608-015-9752-y>
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122. <https://doi.org/10.1006/game.1995.1027>
- Bertsch, K., Hillmann, K., & Herpertz, S. C. (2018). Behavioral and neurobiological correlates of disturbed emotion processing in borderline personality disorder. *Psychopathology, 51*(2), 76–82. <https://doi.org/10.1159/000487363>
- Berwick, D. M., Murphy, J. M., Goldman, P. A., Ware, J. E., Barsky, A. J., & Weinstein, M. C. (1991). Performance of a five-item mental health screening test. *Medical Care, 29*(2), 169–176. <https://doi.org/10.1097/00005650-199102000-00008>

- Bhar, S. S., Brown, G. K., & Beck, A. T. (2008). Dysfunctional beliefs and psychopathology in borderline personality disorder. *Journal of Personality Disorders, 22*(2), 165–177.
<https://doi.org/10.1521/pedi.2008.22.2.165>
- Bhatia, V., Davila, J., Eubanks-Carter, C., & Burckell, L. A. (2013). Appraisals of daily romantic relationship experiences in individuals with borderline personality disorder features. *Journal of Family Psychology, 27*(3), 518–524.
<https://doi.org/10.1037/a0032870>
- Bliese, P. D., Kautz, J., & Lang, J. W. B. (2020). Discontinuous growth models: Illustrations, recommendations, and an R function for generating the design matrix. In Y. Griep & S. D. Hansen (Eds.), *Handbook on the Temporal Dynamics of Organizational Behavior* (pp. 319–350). Edward Elgar Publishing.
<https://doi.org/10.4337/9781788974387.00030>
- Bliese, P. D., & Lang, J. W. B. (2016). Understanding relative and absolute change in discontinuous growth models: Coding alternatives and implications for hypothesis testing. *Organizational Research Methods, 19*, 562–592.
<https://doi.org/10.1177/1094428116633502>
- Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustration. *Organizational Research Methods, 5*(4), 362–387. <https://doi.org/10.1177/109442802237116>
- Bora, E. (2021). A meta-analysis of theory of mind and ‘mentalization’ in borderline personality disorder: A true neuro-social-cognitive or meta-social-cognitive impairment? *Psychological Medicine, 51*(15), 2541–2551.
<http://dx.doi.org/10.1017/S0033291721003718>
- Bouchard, S., Sabourin, S., Lussier, Y., & Villeneuve, E. (2009). Relationship quality and stability in couples when one partner suffers from borderline personality disorder.

Journal of Marital and Family Therapy, 35(4), 446–455.

<https://doi.org/10.1111/j.1752-0606.2009.00151.x>

Bourke, M. E., & Grenyer, B. F. S. (2013). Therapists' accounts of psychotherapy process associated with treating patients with borderline personality disorder. *Journal Of Personality Disorders*, 27(6), 735–745. https://doi.org/10.1521/pedi_2013_27_108

Bowlby, J. (1969). *Attachment and loss*. The Hogarth Press.

Bowlby, J. (1973). *Attachment and loss. Vol. 2: Separation: Anxiety and anger*. The Hogarth Press.

Bowlby, J. (1980). *Attachment and loss. Vol. 3, Loss: Sadness and depression*. The Hogarth Press.

Bowlby, J., Ainsworth, M., Boston, M., & Rosenbluth, D. (1956). The effects of mother-child separation: A follow-up study. *British Journal of Medical Psychology*, 29(3–4), 211–247. <https://doi.org/10.1111/j.2044-8341.1956.tb00915.x>

Brodsky, B. S., Groves, S. A., Oquendo, M. A., Mann, J. J., & Stanley, B. (2006). Interpersonal precipitants and suicide attempts in borderline personality disorder. *Suicide & Life-Threatening Behavior*, 36(3), 313–322. <https://doi.org/10.1521/suli.2006.36.3.313>

Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

Butler, A. C., Brown, G. K., Beck, A. T., & Grisham, J. R. (2002). Assessment of dysfunctional beliefs in borderline personality disorder. *Behaviour Research and Therapy*, 40(10), 1231–1240. [https://doi.org/10.1016/S0005-7967\(02\)00031-1](https://doi.org/10.1016/S0005-7967(02)00031-1)

Camgöz, M. S., & Karapinar, B. P. (2016). Linking secure attachment to commitment: Trust in supervisors. *Leadership and Organization Development Journal*, 37(3), 387–402. <https://doi.org/10.1108/LODJ-07-2014-0130>

- Campbell, C., Tanzer, M., Saunders, R., Booker, T., Allison, E., Li, E., O'Dowda, C., Luyten, P., & Fonagy, P. (2021). Development and validation of a self-report measure of epistemic trust. *PLOS ONE*, *16*(4), e0250264.
<https://doi.org/10.1371/journal.pone.0250264>
- Campbell, L., Simpson, J. A., Boldry, J. G., & Rubin, H. (2010). Trust, variability in relationship evaluations, and relationship processes. *Journal of Personality and Social Psychology*, *99*(1), 14–31. <https://doi.org/10.1037/a0019714>
- Cavelti, M., Thompson, K., Betts, J., Fowler, C., Luebbers, S., Cottton, S. M., & Chanen, A. (2021). Young people with borderline personality disorder have an increased lifetime risk of being the victim of interpersonal violence. *Journal of Interpersonal Violence*.
<https://doi.org/10.1177/0886260520986270>
- Chapman, A. L., Dixon-Gordon, K. L., Butler, S. M., & Walters, K. N. (2015). Emotional reactivity to social rejection versus a frustration induction among persons with borderline personality features. *Personality Disorders: Theory, Research, and Treatment*, *6*(1), 88–96. <https://doi.org/10.1037/per0000101>
- Chiesa, M., & Fonagy, P. (2014). Reflective function as a mediator between childhood adversity, personality disorder and symptom distress. *Personality and Mental Health*, *8*(1), 52–66. <https://doi.org/10.1002/pmh.1245>
- Chita-Tegmark, M. (2018). *Social interactions and the prisoner's dilemma game: New measures of cognitive and behavioral phenotypes* [Ph.D., Boston University].
<https://hdl.handle.net/2144/33085>
- Clarkin, J. F., Yeomans, F. E., & Kernberg, O. F. (1999). *Psychotherapy for borderline personality*. Wiley.

- Clifton, A., Pilkonis, P. A., & McCarty, C. (2007). Social networks in borderline personality disorder. *Journal of Personality Disorders, 21*(4), 434–441.
<https://doi.org/10.1521/pedi.2007.21.4.434>
- Cochard, F., Nguyen Van, P., & Willinger, M. (2004). Trusting behavior in a repeated investment game. *Journal of Economic Behavior & Organization, 55*(1), 31–44.
<https://doi.org/10.1016/j.jebo.2003.07.004>
- Colle, L., Gabbatore, I., Riberi, E., Borroz, E., Bosco, F. M., & Keller, R. (2019). Mindreading abilities and borderline personality disorder: A comprehensive assessment using the Theory of Mind Assessment Scale. *Psychiatry Research, 272*, 609–617. <https://doi.org/10.1016/j.psychres.2018.12.102>
- Collins, N. L., & Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *Journal of Personality and Social Psychology, 58*(4), 644–663. <https://doi.org/10.1037/0022-3514.58.4.644>
- Corgnet, B., Espín, A. M., Hernán-González, R., Kujal, P., & Rassenti, S. (2016). To trust, or not to trust: Cognitive reflection in trust games. *Journal of Behavioral and Experimental Economics, 64*, 20–27. <https://doi.org/10.1016/j.socec.2015.09.008>
- Cortés-García, L., Akça, Ö. F., Wall, K., & Sharp, C. (2021). Exploring mentalizing in adolescents with anorexia nervosa and borderline personality disorder: A comparative study of psychiatric inpatients and healthy controls. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, 9*, 16–26.
<https://doi.org/10.21307/sjcapp-2021-003>
- Daros, A. R., Zakzanis, K. K., & Ruocco, A. C. (2013). Facial emotion recognition in borderline personality disorder. *Psychological Medicine, 43*(9), 1953–1963.
<https://doi.org/10.1017/S0033291712002607>

- De Dreu, C. K. W. (2012). Oxytocin modulates the link between adult attachment and cooperation through reduced betrayal aversion. *Psychoneuroendocrinology*, *37*(7), 871–880. <https://doi.org/10.1016/j.psyneuen.2011.10.003>
- De Meulemeester, C., Lowyck, B., Vermote, R., Verhaest, Y., & Luyten, P. (2017). Mentalizing and interpersonal problems in borderline personality disorder: The mediating role of identity diffusion. *Psychiatry Research*, *258*, 141–144. <https://doi.org/10.1016/j.psychres.2017.09.061>
- De Panfilis, C., Riva, P., Preti, E., Cabrino, C., & Marchesi, C. (2015). When social inclusion is not enough: Implicit expectations of extreme inclusion in borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment*, *6*(4), 301–309. <https://doi.org/10.1037/per0000132>
- De Panfilis, C., Schito, G., Generali, I., Gozzi, L. A., Ossola, P., Marchesi, C., & Grecucci, A. (2019). Emotions at the border: Increased punishment behavior during fair interpersonal exchanges in borderline personality disorder. *Journal of Abnormal Psychology*, *128*(2), 162–172. <https://doi.org/10.1037/abn0000404>
- Declerck, C. H., Boone, C., Pauwels, L., Vogt, B., & Fehr, E. (2020). A registered replication study on oxytocin and trust. *Nature Human Behaviour*, *4*(6), 646–655. <https://doi.org/10.1038/s41562-020-0878-x>
- Derks, J., Van Scheppingen, M. A., Lee, N. C., & Krabbendam, L. (2015). Trust and mindreading in adolescents: The moderating role of social value orientation. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2015.00965>
- DeWall, C. N., MacDonald, G., Webster, G. D., Masten, C. L., Baumeister, R. F., Powell, C., Combs, D., Schurtz, D. R., Stillman, T. F., Tice, D. M., & Eisenberger, N. I. (2010). Acetaminophen Reduces Social Pain: Behavioral and Neural Evidence. *Psychological Science*, *21*(7), 931–937. <https://doi.org/10.1177/0956797610374741>

- Diagnostic and statistical manual of mental disorders: DSM-IV* (4th ed.). (1994). American Psychiatric Association. <https://doi.org/10.1176/ajp.152.8.1228>
- Dimaggio, G., & Lysaker, P. H. (2015). Metacognition and mentalizing in the psychotherapy of patients with psychosis and personality disorders. *Journal of Clinical Psychology, 71*(2), 117–124. <https://doi.org/10.1002/jclp.22147>
- Dimitrijevic, A., Hanak, N., Dimitrijevic, A. A., & Marjanovic, Z. J. (2018). The Mentalization Scale (MentS): A self-report measure for the assessment of mentalizing capacity. *Journal of Personality Assessment, 100*(3), 268–280. <https://doi.org/10.1080/00223891.2017.1310730>
- Dinsdale, N., & Crespi, B. J. (2013). The borderline empathy paradox: Evidence and conceptual models for empathic enhancements in borderline personality disorder. *Journal of Personality Disorders, 27*(2), 172–195. <https://doi.org/10.1521/pedi.2013.27.2.172>
- Dixon-Gordon, K. L., Yiu, A., & Chapman, A. L. (2013). Borderline personality features and emotional reactivity: The mediating role of interpersonal vulnerabilities. *Journal of Behavior Therapy and Experimental Psychiatry, 44*(2), 271–278.
- Domsalla, M., Koppe, G., Niedtfeld, I., Vollstädt-Klein, S., Schmahl, C., Bohus, M., & Lis, S. (2014). Cerebral processing of social rejection in patients with borderline personality disorder. *Social Cognitive and Affective Neuroscience, 9*(11), 1789–1797. <https://doi.org/10.1093/scan/nst176>
- Dougherty, D. M., Bjork, J. M., Huckabee, H. C., Moeller, F. G., & Swann, A. C. (1999). Laboratory measures of aggression and impulsivity in women with borderline personality disorder. *Psychiatry Research, 85*(3), 315–326. [https://doi.org/10.1016/s0165-1781\(99\)00011-6](https://doi.org/10.1016/s0165-1781(99)00011-6)

- Downey, G., & Feldman, S. I. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology*, 70(6), 1327–1343.
<https://doi.org/10.1037//0022-3514.70.6.1327>
- Dreber, A., Fudenberg, D., & Rand, D. G. (2014). Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. *Journal of Economic Behavior & Organization*, 98, 41–55. <https://doi.org/10.1016/j.jebo.2013.12.007>
- Dziobek, I., Preißler, S., Grozdanovic, Z., Heuser, I., Heekeren, H. R., & Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *NeuroImage*, 57(2), 539–548.
<https://doi.org/10.1016/j.neuroimage.2011.05.005>
- Ebert, A., Kolb, M., Heller, J., Edel, M. A., Roser, P., & Brüne, M. (2013). Modulation of interpersonal trust in borderline personality disorder by intranasal oxytocin and childhood trauma. *Social Neuroscience*, 8(4), 305–313.
<https://doi.org/10.1080/17470919.2013.807301>
- Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: A meta-analysis. *Psychotherapy*, 55(4), 508–519. <https://doi.org/10.1037/pst0000185>
- Euler, S., Nolte, T., Constantinou, M., Griem, J., Montague, P. R., & Fonagy, P. (2021). Interpersonal problems in borderline personality disorder: Associations with mentalizing, emotion regulation, and impulsiveness. *Journal of Personality Disorders*, 35(2), 177–193. https://doi.org/10.1521/pedi_2019_33_427
- Eyden, J., Winsper, C., Wolke, D., Broome, M. R., & MacCallum, F. (2016). A systematic review of the parenting and outcomes experienced by offspring of mothers with borderline personality pathology: Potential mechanisms and clinical implications. *Clinical Psychology Review*, 47, 85–105. <https://doi.org/10.1016/j.cpr.2016.04.002>

- Fertuck, E. A., Fischer, S., & Beeney, J. (2018). Social cognition and borderline personality disorder. *Psychiatric Clinics of North America*, *41*(4), 613–632.
<https://doi.org/10.1016/j.psc.2018.07.003>
- Fertuck, E. A., Grinband, J., & Stanley, B. (2013). Facial trust appraisal negatively biased in borderline personality disorder. *Psychiatry Research*, *207*(3), 195–202.
<https://doi.org/10.1016/j.psychres.2013.01.004>
- Fett, A. K. J., Shergill, S. S., Gromann, P. M., Dumontheil, I., Blakemore, S. J., Yakub, F., & Krabbendam, L. (2014). Trust and social reciprocity in adolescence: A matter of perspective-taking. *Journal of Adolescence*, *37*(2), 175–184.
<https://doi.org/10.1016/j.adolescence.2013.11.011>
- Fett, A. K. J., Shergill, S. S., Korver-Nieberg, N., Yakub, F., Gromann, P. M., & Krabbendam, L. (2016). Learning to trust: Trust and attachment in early psychosis. *Psychological Medicine*, *46*(7), 1437–1447.
<https://doi.org/10.1017/S0033291716000015>
- Fitzpatrick, J., & Lafontaine, M.-F. (2017). Attachment, trust, and satisfaction in relationships: Investigating actor, partner, and mediating effects. *Personal Relationships*, *24*(3), 640–662. <https://doi.org/10.1111/pere.12203>
- Flasbeck, V., Enzi, B., & Brune, M. (2017). Altered empathy for psychological and physical pain in borderline personality disorder. *Journal of Personality Disorders*, *31*(5), 689–708. https://doi.org/10.1521/pedi_2017_31_276
- Flasbeck, V., Enzi, B., & Brüne, M. (2019). Childhood trauma affects processing of social interactions in borderline personality disorder: An event-related potential study investigating empathy for pain. *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry*, *20*(4), 278–288. <https://doi.org/10.1080/15622975.2017.1333147>

- Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology*, *21*(4), 1355–1381. <https://doi.org/10.1017/s0954579409990198>
- Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic petrification and the restoration of epistemic trust: A new conceptualization of borderline personality disorder and its psychosocial treatment. *Journal of Personality Disorders*, *29*(5), 575–609. <https://doi.org/10.1521/pedi.2015.29.5.575>
- Fonagy, P., Steele, M., Steele, H., Leigh, T., Kennedy, R., Mattoon, G., & Target, M. (1995). Attachment, the reflective self, and borderline states: The predictive specificity of the Adult Attachment Interview and pathological emotional development. In S. Goldberg, R. Muir, & J. Kerr (Eds.). *Attachment theory: Social, developmental, and clinical perspectives* (pp. 233–278). Analytic Press, Inc.
- Foxhall, M., Hamilton-Giachritsis, C., & Button, K. (2019). The link between rejection sensitivity and borderline personality disorder: A systematic review and meta-analysis. *British Journal of Clinical Psychology*, *58*(3), 289–326. <https://doi.org/10.1111/bjc.12216>
- Fraley, R. C., Hudson, N. W., Heffernan, M. E., & Segal, N. (2015). Are adult attachment styles categorical or dimensional? A taxometric analysis of general and relationship-specific attachment orientations. *Journal of Personality and Social Psychology*, *109*(2), 354–368. <https://doi.org/10.1037/pspp0000027>
- Franzen, N., Hagenhoff, M., Baer, N., Schmidt, A., Mier, D., Sammer, G., Gallhofer, B., Kirsch, P., & Lis, S. (2011). Superior “theory of mind” in borderline personality disorder: An analysis of interaction behavior in a virtual trust game. *Psychiatry Research*, *187*(1–2), 224–233. <https://doi.org/10.1016/j.psychres.2010.11.012>

- Frazier, M. L., Gooty, J., Little, L. M., & Nelson, D. L. (2015). Employee attachment: Implications for supervisor trustworthiness and trust. *Journal of Business and Psychology, 30*(2), 373–386. <https://doi.org/10.1007/s10869-014-9367-4>
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives, 19*(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- Frith, C. D., & Singer, T. (2008). The role of social cognition in decision making. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1511), 3875–3886. <https://doi.org/10.1098/rstb.2008.0156>
- Fulmer, C. A. (2010). *Dynamic trust processes after violation: Trust dissolution and restoration* [University of Maryland]. (Order No. 1529311). Available from ProQuest One Academic. (1170996963).
- Fulmer, C. A., & Gelfand, M. J. (2013). How do I trust thee? Dynamic trust patterns and their individual and social contextual determinants. In K. Sycara, M. Gelfand, & A. Abbe (Eds.), *Models for Intercultural Collaboration and Negotiation* (pp. 97–131). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-5574-1_5
- Fulmer, C. A., & Gelfand, M. J. (2015). Trust after violations: Are collectivists more or less forgiving? *Journal of Trust Research, 5*(2), 109–131. <https://doi.org/10.1080/21515581.2015.1051050>
- Fung, K., & Alden, L. E. (2017). Once hurt, twice shy: Social pain contributes to social anxiety. *Emotion, 17*(2), 231–239. <https://doi.org/10.1037/emo0000223>
- Gadassi, R., Snir, A., Berenson, K., Downey, G., & Rafaeli, E. (2014). Out of the frying pan, into the fire: Mixed affective reactions to social proximity in borderline and avoidant personality disorders in daily life. *Journal of Abnormal Psychology, 123*(3), 613–622. <https://doi.org/10.1037/a0036929>

- Galbally, M., Lewis, A. J., IJzendoorn, M. van, & Permezel, M. (2011). The Role of Oxytocin in Mother-Infant Relations: A Systematic Review of Human Studies. *Harvard Review of Psychiatry, 19*(1), 1–14.
<https://doi.org/10.3109/10673229.2011.549771>
- Germans, S., Van Heck, G. L., Moran, P., & Hodiamont, P. P. G. (2008). The Self-report Standardized Assessment of Personality-abbreviated Scale: Preliminary results of a brief screening test for personality disorders. *Personality and Mental Health, 2*(2), 70–76. <https://doi.org/10.1002/pmh.34>
- Givertz, M., Woszidlo, A., Segrin, C., & Knutson, K. (2013). Direct and indirect effects of attachment orientation on relationship quality and loneliness in married couples. *Journal of Social and Personal Relationships, 30*(8), 1096–1120.
<https://doi.org/10.1177/0265407513482445>
- Graustella, A. J., & MacLeod, C. (2012). A critical review of the influence of oxytocin nasal spray on social cognition in humans: Evidence and future directions. *Hormones and Behavior, 61*(3), 410–418. <https://doi.org/10.1016/j.yhbeh.2012.01.002>
- Grzegorzewski, P., Kulesza, M., Pluta, A., Iqbal, Z., & Kucharska, K. (2019). Assessing self-reported empathy and altruism in patients suffering from enduring borderline personality disorder. *Psychiatry Research, 273*, 798–807.
<https://doi.org/10.1016/j.psychres.2018.12.109>
- Gunderson, J. G. (2007). Disturbed relationships as a phenotype for borderline personality disorder. *The American Journal of Psychiatry, 164*(11), 1637–1640.
<https://doi.org/10.1176/appi.ajp.2007.07071125>
- Gunderson, J. G. (2011). Borderline personality disorder. *New England Journal of Medicine, 364*(21), 2037–2042. <https://doi.org/10.1056/NEJMcp1007358>

- Gunderson, J. G., & Lyons-Ruth, K. (2008). BPD's interpersonal hypersensitivity phenotype: A gene-environment-developmental model. *Journal of Personality Disorders, 22*(1), 22–41. <https://doi.org/10.1521/pedi.2008.22.1.22>
- Gunderson, J. G., Stout, R. L., McGlashan, T. H., Shea, M. T., Morey, L. C., Grilo, C. M., Zanarini, M. C., Yen, S., Markowitz, J. C., Sanislow, C. A., Ansell, E., Pinto, A., & Skodol, A. E. (2011). Ten-year course of borderline personality disorder: Psychopathology and function from the Collaborative Longitudinal Personality Disorders study. *Archives Of General Psychiatry, 68*(8), 827–837. <https://doi.org/10.1001/archgenpsychiatry.2011.37>
- Gutz, L., Renneberg, B., Roepke, S., & Niedeggen, M. (2015). Neural processing of social participation in borderline personality disorder and social anxiety disorder. *Journal of Abnormal Psychology, 124*(2), 421–431. <https://doi.org/10.1037/a0038614>
- Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences, 105*(18), 6741–6746. <https://doi.org/10.1073/pnas.0711099105>
- Harari, H., Shamay-Tsoory, S. G., Ravid, M., & Levkovitz, Y. (2010). Double dissociation between cognitive and affective empathy in borderline personality disorder. *Psychiatry Research, 175*(3), 277–279. <https://doi.org/10.1016/j.psychres.2009.03.002>
- Hepp, J., Lane, S. P., Carpenter, R. W., Niedtfeld, I., Brown, W. C., & Trull, T. J. (2017). Interpersonal problems and negative affect in borderline personality and depressive disorders in daily life. *Clinical Psychological Science, 5*(3), 470–484. <https://doi.org/10.1177/2167702616677312>

- Hepp, J., Lane, S. P., Wycoff, A. M., Carpenter, R. W., & Trull, T. J. (2018). Interpersonal stressors and negative affect in individuals with borderline personality disorder and community adults in daily life: A replication and extension. *Journal of Abnormal Psychology (1965)*, *127*(2), 183–189. <https://doi.org/10.1037/abn0000318>
- Hepp, J., Störkel, L. M., Kieslich, P. J., Schmahl, C., & Niedtfeld, I. (2018). Negative evaluation of individuals with borderline personality disorder at zero acquaintance. *Behaviour Research and Therapy*, *111*, 84–91. <https://doi.org/10.1016/j.brat.2018.09.009>
- Holm, A. L., & Severinsson, E. (2011). Struggling to recover by changing suicidal behaviour: Narratives from women with borderline personality disorder. *International Journal of Mental Health Nursing*, *20*(3), 165–173. <https://doi.org/10.1111/j.1447-0349.2010.00713.x>
- Holwerda, N., Sanderman, R., Pool, G., Hinnen, C., Langendijk, J. A., Bemelman, W. A., Hagedoorn, M., & Sprangers, M. A. G. (2013). Do patients trust their physician? The role of attachment style in the patient-physician relationship within one year after a cancer diagnosis. *Acta Oncologica*, *52*(1), 110–117. <https://doi.org/10.3109/0284186X.2012.689856>
- Homan, P., Reddan, M. C., Brosch, T., Koenigsberg, H. W., & Schiller, D. (2017). Aberrant link between empathy and social attribution style in borderline personality disorder. *Journal of Psychiatric Research*, *94*, 163–171. <https://doi.org/10.1016/j.jpsychires.2017.07.012>
- Howard, K. P., Lazarus, S. A., & Cheavens, J. S. (2022). A longitudinal examination of the reciprocal relationship between borderline personality features and interpersonal relationship quality. *Personality Disorders*, *13*(1), 3–11. <https://doi.org/10.1037/per0000484>

- Hula, A., Vilares, I., Lohrenz, T., Dayan, P., & Montague, P. R. (2018). A model of risk and mental state shifts during social interaction. *PLoS Computational Biology*, *14*(2). Scopus. <https://doi.org/10.1371/journal.pcbi.1005935>
- Jeung, H., & Herpertz, S. C. (2014). Impairments of interpersonal functioning: Empathy and intimacy in borderline personality disorder. *Psychopathology*, *47*(4), 220–234. <https://doi.org/10.1159/000357191>
- Jeung, H., Schwieren, C., & Herpertz, S. C. (2016). Rationality and self-interest as economic-exchange strategy in borderline personality disorder: Game theory, social preferences, and interpersonal behavior. *Neuroscience and Biobehavioral Reviews*, *71*, 849–864. <https://doi.org/10.1016/j.neubiorev.2016.10.030>
- Jeung, H., Vollmann, M., Herpertz, S. C., & Schwieren, C. (2020). Consider others better than yourself: Social decision-making and partner preference in Borderline Personality Disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *67*, 101436. <https://doi.org/10.1016/j.jbtep.2018.11.004>
- Joel, S., Macdonald, G., & Page-Gould, E. (2017). Wanting to stay and wanting to go: Unpacking the content and structure of relationship stay/leave decision processes. *Social Psychological and Personality Science*, *9*(6), 631–644. <https://doi.org/10.1177/1948550617722834>
- Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. <https://doi.org/10.1016/j.joep.2011.05.007>
- Kaletsch, M., Krüger, B., Pilgramm, S., Stark, R., Lis, S., Gallhofer, B., Zentgraf, K., Munzert, J., & Sammer, G. (2014). Borderline personality disorder is associated with lower confidence in perception of emotional body movements. *Frontiers in Psychology*, *5*(NOV), 1262. Scopus. <https://doi.org/10.3389/fpsyg.2014.01262>

- Karantzas, G. C., Feeney, J. A., Goncalves, C. V., & McCabe, M. P. (2014). Towards an integrative attachment-based model of relationship functioning. *British Journal of Psychology, 105*(3), 413–434. <https://doi.org/10.1111/bjop.12047>
- Katsakou, C., Marougka, S., Barnicot, K., Savill, M., White, H., Lockwood, K., & Priebe, S. (2012). Recovery in borderline personality disorder (BPD): A qualitative study of service users' perspectives. *PLOS ONE, 7*(5), e36517. <https://doi.org/10.1371/journal.pone.0036517>
- Kaurin, A., Dombrowski, A. Y., Hallquist, M. N., & Wright, A. G. C. (2020). Momentary interpersonal processes of suicidal surges in borderline personality disorder. *Psychological Medicine, 1–11*. <https://doi.org/10.1017/S0033291720004791>
- Keelan, J. P. R., Dion, K. L., & Dion, K. K. (1994). Attachment style and heterosexual relationships among young adults: A short-term panel study. *Journal of Social and Personal Relationships, 11*(2), 201–214. <https://doi.org/10.1177/0265407594112003>
- Kernberg, O. (1968). The treatment of patients with borderline personality organization. *The International Journal of Psycho-Analysis, 49*(4), 600–619.
- King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, R. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science, 321*(5890), 806–810. <https://doi.org/10.1126/science.1156902>
- Kishida, K. T., King-Casas, B., & Montague, P. R. (2010). Neuroeconomic approaches to mental disorders. *Neuron, 67*(4), 543–554. <https://doi.org/10.1016/j.neuron.2010.07.021>
- Klest, B., & Philippon, O. (2016). Trust in the medical profession and patient attachment style. *Psychology, Health and Medicine, 21*(7), 863–870. <https://doi.org/10.1080/13548506.2015.1120328>

- Korsgaard, M. A., Kautz, J., Bliese, P., Samson, K., & Kostyszyn, P. (2018). Conceptualising time as a level of analysis: New directions in the analysis of trust dynamics. *Journal of Trust Research*, 8(2), 142–165. <https://doi.org/10.1080/21515581.2018.1516557>
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, 435(7042), 673–676. <https://doi.org/10.1038/nature03701>
- Kuhlken, K., Robertson, C., Benson, J., & Nelson-Gray, R. (2014). The interaction of borderline personality disorder symptoms and relationship satisfaction in predicting affect. *Personality Disorders*, 5(1), 20–25. <https://doi.org/10.1037/per0000013>
- Kvarstein, E. H., Pedersen, G., Urnes, O., Hummelen, B., Wilberg, T., & Karterud, S. (2015). Changing from a traditional psychodynamic treatment programme to mentalization-based treatment for patients with borderline personality disorder—Does it make a difference? *Psychology and Psychotherapy-Theory Research and Practice*, 88(1), 71–86. <https://doi.org/10.1111/papt.12036>
- Langley, G. C., & Klopper, H. (2005). Trust as a foundation for the therapeutic intervention for patients with borderline personality disorder. *Journal of Psychiatric and Mental Health Nursing*, 12(1), 23–32. <https://doi.org/10.1111/j.1365-2850.2004.00774.x>
- Lazarus, S. A., Beeney, J. E., Howard, K. P., Strunk, D. R., Pilkonis, P. A., & Cheavens, J. S. (2020). Characterization of relationship instability in women with borderline personality disorder: A social network analysis. *Personality Disorders*, 11(5), 312–320. <https://doi.org/10.1037/per0000380>
- Lazarus, S. A., & Cheavens, J. S. (2017). An examination of social network quality and composition in women with and without borderline personality disorder. *Personality Disorders-Theory Research and Treatment*, 8(4), 340–348. <https://doi.org/10.1037/per0000201>

- Lazarus, S. A., Cheavens, J. S., Festa, F., & Zachary Rosenthal, M. (2014). Interpersonal functioning in borderline personality disorder: A systematic review of behavioral and laboratory-based assessments. *Clinical Psychology Review, 34*(3), 193–205.
<https://doi.org/10.1016/j.cpr.2014.01.007>
- Lazarus, S. A., Choukas-Bradley, S., Beeney, J. E., Byrd, A. L., Vine, V., & Stepp, S. D. (2019). Too much too soon?: Borderline personality disorder symptoms and romantic relationships in adolescent girls. *Journal of Abnormal Child Psychology, 47*(12), 1995–2005. <https://doi.org/10.1007/s10802-019-00570-1>
- Lazarus, S. A., Scott, L. N., Beeney, J. E., Wright, A. G. C., Stepp, S. D., & Pilkonis, P. A. (2018). Borderline personality disorder symptoms and affective responding to perceptions of rejection and acceptance from romantic versus nonromantic partners. *Personality Disorders: Theory, Research, and Treatment, 9*(3), 197–206.
<https://doi.org/10.1037/per0000289>
- Lazarus, S. A., Southward, M. W., & Cheavens, J. S. (2016). Do borderline personality disorder features and rejection sensitivity predict social network outcomes over time? *Personality and Individual Differences, 100*, 62–67.
<https://doi.org/10.1016/j.paid.2016.02.032>
- Lee, V. K., & Harris, L. T. (2013). How social cognition can inform social decision making. *Frontiers in Neuroscience, 7*, 13. <https://doi.org/10.3389/fnins.2013.00259>
- Lévay, E. E., Bajzát, B., & Unoka, Z. S. (2021). Expectation of selfishness from others in borderline personality disorder. *Frontiers in Psychology, 12*, 702227.
<https://doi.org/10.3389/fpsyg.2021.702227>
- Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions.

Journal of Management, 32(6), 991–1022.

<https://doi.org/10.1177/0149206306294405>

- Liebke, L., Koppe, G., Bungert, M., Thome, J., Hauschild, S., Defiebre, N., Izurieta Hidalgo, N. A., Schmahl, C., Bohus, M., & Lis, S. (2018). Difficulties with being socially accepted: An experimental study in borderline personality disorder. *Journal of Abnormal Psychology*. <https://doi.org/10.1037/abn0000373>
- Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder* (pp. xvii, 558). Guilford Press.
- Links, P. S., & Heslegrave, R. J. (2000). Prospective studies of outcome: Understand mechanisms of change in patients with borderline personality disorder. *Psychiatric Clinics of North America*, 23(1), 137–150. [https://doi.org/10.1016/S0193-953X\(05\)70148-9](https://doi.org/10.1016/S0193-953X(05)70148-9)
- Lis, S., & Bohus, M. (2013). Social interaction in borderline personality disorder. *Current Psychiatry Reports*, 15(2). <https://doi.org/10.1007/s11920-012-0338-z>
- Lobbestael, J., & McNally, R. J. (2016). An empirical test of rejection- and anger-related interpretation bias in borderline personality disorder. *Journal of Personality Disorders*, 30(3), 307–319. https://doi.org/10.1521/pedi_2015_29_194
- Lount, R. B., Jr., Zhong, C. B., Sivanathan, N., & Murnighan, J. K. (2008). Getting off on the wrong foot: The timing of a breach and the restoration of trust. *Personality & Social Psychology Bulletin*, 34(12), 1601–1612. <https://doi.org/10.1177/0146167208324512>
- MacDonald, K., & MacDonald, T. M. (2010). The peptide that binds: A systematic review of Oxytocin and its prosocial effects in humans. *Harvard Review of Psychiatry*, 18(1), 1–21. <https://doi.org/10.3109/10673220903523615>

- Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs of the Society for Research in Child Development*, 50(1/2), 66–104. <https://doi.org/10.2307/3333827>
- Malda-Castillo, J., Browne, C., & Perez-Algorta, G. (2018). Mentalization-based treatment and its evidence-base status: A systematic literature review. *Psychology and Psychotherapy: Theory, Research and Practice*. <https://doi.org/10.1111/papt.12195>
- Maneta, E. K., Cohen, S., Schulz, M. S., & Waldinger, R. J. (2013). Two to tango: A dyadic analysis of links between borderline personality traits and intimate partner violence. *Journal of Personality Disorders*, 27(2), 233–243. https://doi.org/10.1521/pedi_2013_27_082
- Martin, F., Flasbeck, V., Brown, E. C., & Brüne, M. (2017). Altered mu-rhythm suppression in borderline personality disorder. *Brain Research*, 1659, 64–70. <https://doi.org/10.1016/j.brainres.2017.01.023>
- Masland, S. R., & Hooley, J. M. (2020). When trust does not come easily: Negative emotional information unduly influences trustworthiness appraisals for individuals with borderline personality features. *Journal of Personality Disorders*, 34(3), 394–409. https://doi.org/10.1521/pedi_2019_33_404
- Masland, S. R., Schnell, S. E., & Shah, T. V. (2020). Trust beliefs, biases, and behaviors in borderline personality disorder: Empirical findings and relevance to epistemic trust. *Current Behavioral Neuroscience Reports*, 7(4), 239–249. <https://doi.org/10.1007/s40473-020-00220-7>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.2307/258792>

- McCloskey, M. S., New, A. S., Siever, L. J., Goodman, M., Koenigsberg, H. W., Flory, J. D., & Coccaro, E. F. (2009). Evaluation of behavioral impulsivity and aggression tasks as endophenotypes for borderline personality disorder. *Journal of Psychiatric Research, 43*(12), 1036–1048. <https://doi.org/10.1016/j.jpsychires.2009.01.002>
- McClure, G., Hawes, D. J., & Dadds, M. R. (2016). Borderline personality disorder and neuropsychological measures of executive function: A systematic review. *Personality and Mental Health, 10*(1), 43–57. <https://doi.org/10.1002/pmh.1320>
- McClure, M. J., Bartz, J. A., & Lydon, J. E. (2013). Uncovering and overcoming ambivalence: The role of chronic and contextually activated attachment in two-person social dilemmas. *Journal of Personality, 81*(1), 103–117. <https://doi.org/10.1111/j.1467-6494.2012.00788.x>
- McLaren, V., Gallagher, M., Hopwood, C. J., & Sharp, C. (2022). Hypermentalizing and Borderline Personality Disorder: A Meta-Analytic Review. *American Journal of Psychotherapy, 75*(1), 21–31. <https://doi.org/10.1176/appi.psychotherapy.20210018>
- Miano, A., Dziobek, I., & Roepke, S. (2017). Understanding interpersonal dysfunction in borderline personality disorder: A naturalistic dyadic study reveals absence of relationship-protective empathic inaccuracy. *Clinical Psychological Science, 5*(2), 355–366. <https://doi.org/10.1177/2167702616683505>
- Miano, A., Fertuck, E. A., Arntz, A., & Stanley, B. (2013). Rejection sensitivity is a mediator between borderline personality disorder features and facial trust appraisal. *Journal of Personality Disorders, 27*(4), 442–456. https://doi.org/10.1521/pedi_2013_27_096
- Miano, A., Fertuck, E. A., Roepke, S., & Dziobek, I. (2017). Romantic relationship dysfunction in borderline personality disorder: A naturalistic approach to trustworthiness perception. *Personality Disorders: Theory, Research, and Treatment, 8*(3), 281–286. <https://doi.org/10.1037/per0000196>

- Mikulincer, M. (1998). Attachment working models and the sense of trust: An exploration of interaction goals and affect regulation. *Journal of Personality and Social Psychology*, 74(5), 1209–1224. <https://doi.org/10.1037/0022-3514.74.5.1209>
- Minzenberg, M. J., Poole, J. H., & Vinogradov, S. (2006). Social-emotion recognition in borderline personality disorder. *Comprehensive Psychiatry*, 47(6), 468–474. <https://doi.org/10.1016/j.comppsy.2006.03.005>
- Mitchell, A. E., Dickens, G. L., & Picchioni, M. M. (2014). Facial emotion processing in borderline personality disorder: A systematic review and meta-analysis. *Neuropsychology Review*, 24(2), 166–184. <https://doi.org/10.1007/s11065-014-9254-9>
- Murnighan, J. K. (1981). Defectors, vulnerability, and relative power: Some causes and effects of leaving a stable coalition. *Human Relations*, 34(7), 589–609. <https://doi.org/10.1177/001872678103400706>
- Navarro-Gomez, S., Frias, A., & Palma, C. (2017). Romantic relationships of people with borderline personality: A narrative review. *Psychopathology*, 50(3), 175–187. <https://doi.org/10.1159/000474950>
- Németh, N., Mátrai, P., Hegyi, P., Czéh, B., Czopf, L., Hussain, A., Pammer, J., Szabó, I., Solymár, M., Kiss, L., Hartmann, P., Szilágyi, Á. L., Kiss, Z., & Simon, M. (2018). Theory of mind disturbances in borderline personality disorder: A meta-analysis. *Psychiatry Research*, 270, 143–153. <https://doi.org/10.1016/j.psychres.2018.08.049>
- New, A. S., Aan Het Rot, M., Ripoll, L. H., Perez-Rodriguez, M. M., Lazarus, S., Zipursky, E., Weinstein, S. R., Koenigsberg, H. W., Hazlett, E. A., Goodman, M., & Siever, L. J. (2012). Empathy and alexithymia in borderline personality disorder: Clinical and laboratory measures. *Journal of Personality Disorders*, 26(5), 660–675. <https://doi.org/10.1521/pedi.2012.26.5.660>

- New, A. S., Hazlett, E. A., Newmark, R. E., Zhang, J., Triebwasser, J., Meyerson, D., Lazarus, S., Trisdorfer, R., Goldstein, K. E., Goodman, M., Koenigsberg, H. W., Flory, J. D., Siever, L. J., & Buchsbaum, M. S. (2009). Laboratory induced aggression: A positron emission tomography study of aggressive individuals with borderline personality disorder. *Biological Psychiatry*, *66*(12), 1107–1114.
<https://doi.org/10.1016/j.biopsych.2009.07.015>
- Ng, F. Y. Y., Bourke, M. E., & Grenyer, B. F. S. (2016). Recovery from borderline personality disorder: A systematic review of the perspectives of consumers, clinicians, family and carers. *PLOS ONE*, *11*(8), e0160515.
<https://doi.org/10.1371/journal.pone.0160515>
- Nica, E. I., & Links, P. S. (2009). Affective instability in borderline personality disorder: Experience sampling findings. *Current Psychiatry Reports*, *11*(1), 74–81.
<https://doi.org/10.1007/s11920-009-0012-2>
- Nicol, K., Pope, M., Sprengelmeyer, R., Young, A. W., & Hall, J. (2013). Social judgement in borderline personality disorder. *PLoS ONE*, *8*(11).
<https://doi.org/10.1371/journal.pone.0073440>
- Niedtfeld, I., & Kroneisen, M. (2020). Impaired memory for cooperative interaction partners in borderline personality disorder. *Borderline Personality Disorder and Emotion Dysregulation*, *7*(1), 22. <https://doi.org/10.1186/s40479-020-00137-3>
- Niessen, C., & Jimmieson, N. L. (2016). Threat of resource loss: The role of self-regulation in adaptive task performance. *The Journal of Applied Psychology*, *101*(3), 450–462.
<https://doi.org/10.1037/apl0000049>
- Normann-Eide, E., Antonsen, B. mar T., Kvarstein, E. H., Pedersen, G., Vaskinn, A., & Wilberg, T. (2020). Are impairments in theory of mind specific to borderline

- personality disorder? *Journal of Personality Disorders*, 34(6), 827–841.
https://doi.org/10.1521/pedi_2019_33_417
- Orme, W., Bowersox, L., Vanwoerden, S., Fonagy, P., & Sharp, C. (2019). The relation between epistemic trust and borderline pathology in an adolescent inpatient sample. *Borderline Personality Disorder and Emotion Dysregulation*, 6(1), 13.
<https://doi.org/10.1186/s40479-019-0110-7>
- Paret, C., Jennen-Steinmetz, C., & Schmahl, C. (2017). Disadvantageous decision-making in borderline personality disorder: Partial support from a meta-analytic review. *Neuroscience and Biobehavioral Reviews*, 72, 301–309.
<https://doi.org/10.1016/j.neubiorev.2016.11.019>
- Penner, F., McLaren, V., Leavitt, J., Akca, O. F., & Sharp, C. (2020). Implicit and explicit mentalizing deficits in adolescent inpatients: Specificity and incremental value of borderline pathology. *Journal of Personality Disorders*, 34(Supplement B), 64–83.
https://doi.org/10.1521/pedi_2019_33_463
- Perroud, N., Badoud, D., Weibel, S., Nicastro, R., Hasler, R., Kung, A. L., Luyten, P., Fonagy, P., Dayer, A., Aubry, J. M., Prada, P., & Debbane, M. (2017). Mentalization in adults with attention deficit hyperactivity disorder: Comparison with controls and patients with borderline personality disorder. *Psychiatry Research*, 256, 334–341.
<https://doi.org/10.1016/j.psychres.2017.06.087>
- Petersen, R., Brakoulias, V., & Langdon, R. (2016). An experimental investigation of mentalization ability in borderline personality disorder. *Comprehensive Psychiatry*, 64, 12–21. <https://doi.org/10.1016/j.comppsy.2015.10.004>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. D. (2019). *NLME: Linear and nonlinear mixed effects models*. <https://svn.r-project.org/R-packages/trunk/nlme>

- Pluta, A., Kulesza, M., Grzegorzewski, P., & Kucharska, K. (2018). Assessing advanced theory of mind and alexithymia in patients suffering from enduring borderline personality disorder. *Psychiatry Research, 261*, 436–441. <https://doi.org/10.1016/j.psychres.2018.01.003>
- Poggi, A., Richetin, J., & Preti, E. (2019). Trust and rejection sensitivity in personality disorders. *Current Psychiatry Reports, 21*(8), 69. <https://doi.org/10.1007/s11920-019-1059-3>
- Polgar, P., Fogd, D., Unoka, Z., Siraly, E., & Csukly, G. (2014). Altered social decision making in borderline personality disorder: An ultimatum game study. *Journal of Personality Disorders, 28*(6), 841–852. https://doi.org/10.1521/pedi_2014_28_142
- Preißler, S., Dziobek, I., Ritter, K., Heekeren, H. R., & Roepke, S. (2010). Social cognition in borderline personality disorder: Evidence for disturbed recognition of the emotions, thoughts, and intentions of others. *Frontiers in Behavioral Neuroscience, 4*(DEC), 182. <https://doi.org/10.3389/fnbeh.2010.00182>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral & Brain Sciences, 1*(4), 515. <https://doi.org/10.1017/S0140525X00076512>
- Quek, J., Melvin, G. A., Bennett, C., Gordon, M. S., Saeedi, N., & Newman, L. K. (2019). Mentalization in adolescents with borderline personality disorder: A comparison with healthy controls. *Journal of Personality Disorders, 33*(2), 145–163. https://doi.org/10.1521/pedi_2018_32_336
- R Development Core Team. (2018). *R: A Language and Environment for Statistical Computing* (Version 3.6.1). R Foundation for Statistical Computing. <https://www.R-project.org>
- Reichenberger, J., Eibl, J. J., Pfaltz, M., Wilhelm, F. H., Voderholzer, U., Hillert, A., & Blechert, J. (2017). Don't praise me, don't chase me: Emotional reactivity to positive

and negative social-evaluative videos in patients with borderline personality disorder. *Journal of Personality Disorders*, 31(1), 75–89.

https://doi.org/10.1521/pedi_2016_30_238

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, 49(1), 95–112. <https://doi.org/10.1037/0022-3514.49.1.95>

Renneberg, B., Herm, K., Hahn, A., Staebler, K., Lammers, C. H., & Roepke, S. (2012). Perception of social participation in borderline personality disorder. *Clinical Psychology and Psychotherapy*, 19(6), 473–480. <https://doi.org/10.1002/cpp.772>

Reuter, T. R., Sharp, C., Temple, J. R., & Babcock, J. C. (2015). The relation between borderline personality disorder features and teen dating violence. *Psychology of Violence*, 5(2), 163–173. <https://doi.org/10.1037/a0037891>

Richetin, J., Poggi, A., Ricciardelli, P., Fertuck, E. A., & Preti, E. (2018). The emotional components of rejection sensitivity as a mediator between borderline personality disorder and biased appraisal of trust in faces. *Clinical Neuropsychiatry*, 15(4), 200–205.

Richman, M. J., & Unoka, Z. (2015). Mental state decoding impairment in major depression and borderline personality disorder: Meta-analysis. *British Journal of Psychiatry*, 207(6), 483–489. <https://doi.org/10.1192/bjp.bp.114.152108>

Richmond, J. R., Edmonds, K. A., Rose, J. P., & Gratz, K. L. (2020). Examining the impact of online rejection among emerging adults with borderline personality pathology: Development of a novel online group chat social rejection paradigm. *Personality Disorders*, 11(5), 301–311. <https://doi.org/10.1037/per0000381>

- Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, 62, 23–48.
<https://doi.org/10.1146/annurev.psych.121208.131647>
- Ritter, K., Dziobek, I., Preißler, S., Rüter, A., Vater, A., Fydrich, T., Lammers, C. H., Heekeren, H. R., & Roepke, S. (2011). Lack of empathy in patients with narcissistic personality disorder. *Psychiatry Research*, 187(1–2), 241–247.
<https://doi.org/10.1016/j.psychres.2010.09.013>
- Roberts, I. D., Krajchich, I., Cheavens, J. S., Campo, J. V., & Way, B. M. (2018). Acetaminophen reduces distrust in individuals with borderline personality disorder features. *Clinical Psychological Science*, 6(1), 145–154.
<https://doi.org/10.1177/2167702617731374>
- Roepke, S., Vater, A., Preißler, S., Heekeren, H. R., & Dziobek, I. (2012). Social cognition in borderline personality disorder. *Frontiers in Neuroscience*, JAN.
<https://doi.org/10.3389/fnins.2012.00195>
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393–404.
<https://doi.org/10.5465/amr.1998.926617>
- Runions, K. C., Wong, J., Pace, G., & Salmin, I. (2021). Borderline personality disorder and peers: A scoping review of friendship, victimization and aggression studies. *Adolescent Research Review*, 6(4), 359–389. <https://doi.org/10.1007/s40894-020-00137-y>
- Ruocco, A. C. (2005). The neuropsychology of borderline personality disorder: A meta-analysis and review. *Psychiatry Research*, 137(3), 191–202.
<https://doi.org/10.1016/j.psychres.2005.07.004>

- Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition*, 55(1), 209–219.
<https://doi.org/10.1016/j.bandc.2003.04.002>
- Sadikaj, G., Russell, J. J., Moskowitz, D. S., & Paris, J. (2010). Affect dysregulation in individuals with borderline personality disorder: Persistence and interpersonal triggers. *Journal of Personality Assessment*, 92(6), 490–500.
<https://doi.org/10.1080/00223891.2010.513287>
- Safran, J. D. (1993). Breaches in the therapeutic alliance: An arena for negotiating authentic relatedness. *Psychotherapy: Theory, Research, Practice, Training*, 30(1), 11–24.
<https://doi.org/10.1037/0033-3204.30.1.11>
- Safran, J. D., Muran, J. C., & Eubanks-Carter, C. (2011). Repairing alliance ruptures. *Psychotherapy*, 48(1), 80–87. <https://doi.org/10.1037/a0022140>
- Salgado, R. M., Pedrosa, R., & Bastos-Leite, A. J. (2020). Dysfunction of empathy and related processes in borderline personality disorder: A systematic review. *Harvard Review of Psychiatry*, 28(4), 238–254.
<https://doi.org/10.1097/HRP.0000000000000260>
- Sanislow, C. A., Grilo, C. M., & McGlashan, T. H. (2000). Factor analysis of the DSM-III-R borderline personality disorder criteria in psychiatric inpatients. *The American Journal of Psychiatry*, 157(10), 1629–1633.
<https://doi.org/10.1176/appi.ajp.157.10.1629>
- Sanislow, C. A., Grilo, C. M., Morey, L. C., Bender, D. S., Skodol, A. E., Gunderson, J. G., Shea, M. T., Stout, R. L., Zanarini, M. C., & McGlashan, T. H. (2002). Confirmatory factor analysis of DSM-IV criteria for borderline personality disorder: Findings from the collaborative longitudinal personality disorders study. *The American Journal of Psychiatry*, 159(2), 284–290. <https://doi.org/10.1176/appi.ajp.159.2.284>

- Sansone, R. A., & Sansone, L. A. (2011). Personality disorders: A nation-based perspective on prevalence. *Innovations in Clinical Neuroscience*, 8(4), 13–18.
- Saunders, K. E. A., Goodwin, G. M., & Rogers, R. D. (2015). Borderline personality disorder, but not euthymic bipolar disorder, is associated with a failure to sustain reciprocal cooperative behaviour: Implications for spectrum models of mood disorders. *Psychological Medicine*, 45(8), 1591–1600.
<https://doi.org/10.1017/S0033291714002475>
- Schenk, N., Fürer, L., Zimmermann, R., Steppan, M., & Schmeck, K. (2020). Alliance ruptures and resolutions in personality disorders. *Current Psychiatry Reports*, 23(1), 1. <https://doi.org/10.1007/s11920-020-01212-w>
- Schilke, O., Reimann, M., & Cook, K. S. (2013). Effect of relationship experience on trust recovery following a breach. *Proceedings of the National Academy of Sciences*, 110(38), 15236–15241. <https://doi.org/10.1073/pnas.131485711>
- Schilling, L., Wingenfeld, K., Löwe, B., Moritz, S., Terfehr, K., Köther, U., & Spitzer, C. (2012). Normal mind-reading capacity but higher response confidence in borderline personality disorder patients. *Psychiatry and Clinical Neurosciences*, 66(4), 322–327. <https://doi.org/10.1111/j.1440-1819.2012.02334.x>
- Schneider, I., Boll, S., Volman, I., Roelofs, K., Spohn, A., Herpertz, S. C., & Bertsch, K. (2020). Oxytocin normalizes approach–avoidance behavior in women with borderline personality disorder. *Frontiers in Psychiatry*, 11.
<https://www.frontiersin.org/articles/10.3389/fpsy.2020.00120>
- Scott, L. N., Wright, A. G. C., Beeney, J. E., Lazarus, S. A., Pilkonis, P. A., & Stepp, S. D. (2017). Borderline personality disorder symptoms and aggression: A within-person process model. *Journal of Abnormal Psychology*, 126(4), 429–440.
<https://doi.org/10.1037/abn0000272>

- Segal, H. G., Westen, D., Lohr, N. E., R. Silk, K., & Cohen, R. (1992). Assessing object relations and social cognition in borderline personality disorders from stories told to the Picture Arrangement subtest of the WAIS-R. *Journal of Personality Disorders*, 6(4), 458–470. <https://doi.org/10.1521/pedi.1992.6.4.458>
- Segal, H. G., Westen, D., Lohr, N. E., & Silk, K. R. (1993). Clinical assessment of object relations and social cognition using stories told to the Picture Arrangement subtest of the WAIS-R. *Journal of Personality Assessment*, 61(1), 58–80. https://doi.org/10.1207/s15327752jpa6101_5
- Semerari, A., Carcione, A., Dimaggio, G., Falcone, M., Nicolò, G., Procacci, M., & Alleva, G. (2003). How to evaluate metacognitive functioning in psychotherapy? The metacognition assessment scale and its applications. *Clinical Psychology & Psychotherapy*, 10(4), 238–261. <https://doi.org/10.1002/cpp.362>
- Sharp, C. (2014). The social-cognitive basis of BPD: A theory of hypermentalizing. In C. Sharp & J. L. Tackett (Eds.), *Handbook of borderline personality disorder in children and adolescents* (pp. 211–225). Springer. https://doi.org/10.1007/978-1-4939-0591-1_15
- Sharp, C., Fonagy, P., & Goodyer, I. (Eds.). (2008). Social cognition and developmental psychopathology. In *Social Cognition and Developmental Psychopathology*. Oxford University Press. 10.1093/med/9780198569183.001.0001
- Sharp, C., Ha, C., Carbone, C., Kim, S., Perry, K., Williams, L., & Fonagy, P. (2013). Hypermentalizing in adolescent inpatients: Treatment effects and association with borderline traits. *Journal of Personality Disorders*, 27(SPL.ISS.1), 3–18. <https://doi.org/10.1521/pedi.2013.27.1.3>
- Sharp, C., Ha, C., & Fonagy, P. (2011). Get them before they get you: Trust, trustworthiness, and social cognition in boys with and without externalizing behavior problems.

- Development and Psychopathology*, 23(2), 647–658. Scopus.
<https://doi.org/10.1017/S0954579410000003>
- Sharp, C., Pane, H., Ha, C., Venta, A., Patel, A. B., Sturek, J., & Fonagy, P. (2011). Theory of mind and emotion regulation difficulties in adolescents with borderline traits. *Journal of the American Academy of Child and Adolescent Psychiatry*, 50(6), 563–573. <https://doi.org/10.1016/j.jaac.2011.01.017>
- Sharp, C., & Vanwoerden, S. (2015). Hypermentalizing in borderline personality disorder: A model and data. *Journal of Infant, Child, and Adolescent Psychotherapy*, 14(1), 33–45. <https://doi.org/10.1080/15289168.2015.1004890>
- Simpson, J. A. (1990). Influence of attachment styles on romantic relationships. *Journal of Personality and Social Psychology*, 59(5), 971–980. <https://doi.org/10.1037/0022-3514.59.5.971>
- Simpson, J. A. (2007). Foundations of interpersonal trust. In *Social psychology: Handbook of basic principles, 2nd ed* (pp. 587–607). The Guilford Press.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780195152968.001.0001>
- Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic “remedies” for trust/distrust. *Organization Science*, 4(3), 367–392.
<https://doi.org/10.1287/orsc.4.3.367>
- Snijders, J. D., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. SAGE.
- Somma, A., Ferrara, M., Terrinoni, A., Frau, C., Ardizzone, I., Sharp, C., & Fossati, A. (2019). Hypermentalizing as a marker of borderline personality disorder in Italian adolescents: A cross-cultural replication of Sharp and colleagues’ (2011) findings.

Borderline Personality Disorder and Emotion Dysregulation, 6(1), 1–4.

<https://doi.org/10.1186/s40479-019-0104-5>

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D.

(2010). Epistemic Vigilance. *Mind & Language*, 25(4), 359–393.

<https://doi.org/10.1111/j.1468-0017.2010.01394.x>

Staebler, K., Helbing, E., Rosenbach, C., & Renneberg, B. (2011). Rejection sensitivity and

borderline personality disorder. *Clinical Psychology and Psychotherapy*, 18(4), 275–

283. <https://doi.org/10.1002/cpp.705>

Stepp, S. D., Pilkonis, P. A., Yaggi, K. E., Morse, J. Q., & Feske, U. (2009). Interpersonal

and emotional experiences of social interactions in borderline personality disorder.

The Journal of Nervous and Mental Disease, 197(7), 484–491.

<https://doi.org/10.1097/NMD.0b013e3181aad2e7>

Stiglmayr, C. E., Grathwol, T., Linehan, M. M., Ihorst, G., Fahrenberg, J., & Bohus, M.

(2005). Aversive tension in patients with borderline personality disorder: A computer-based controlled field study. *Acta Psychiatrica Scandinavica*, 111(5), 372–379.

<https://doi.org/10.1111/j.1600-0447.2004.00466.x>

Sun, B., Yu, X., Yuan, X., Sun, C., & Li, W. (2021). The effect of social perspective-taking

on interpersonal trust under the cooperative and competitive contexts: The mediating role of benevolence. *Psychology Research and Behavior Management*, 14, 817–826.

<https://doi.org/10.2147/PRBM.S310557>

Szymanska, M., Schneider, M., Chateau-Smith, C., Nezelof, S., & Vulliez-Coady, L. (2017).

Psychophysiological effects of oxytocin on parent-child interactions: A literature review on oxytocin and parent-child interactions. *Psychiatry and Clinical*

Neurosciences, 71(10), 690–705. <https://doi.org/10.1111/pcn.12544>

- Taheri, M., Rotshtein, P., & Beierholm, U. (2018). The effect of attachment and environmental manipulations on cooperative behavior in the prisoner's dilemma game. *PLOS ONE*, *13*(11), e0205730. <https://doi.org/10.1371/journal.pone.0205730>
- Thayer, A. (2015). *Trust trajectories as a function of violation type and repair efforts* [Electronic Theses and Dissertations]. <https://stars.library.ucf.edu/etd/1252>
- Theodoridou, A., Rowe, A. C., Penton-Voak, I. S., & Rogers, P. J. (2009). Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness. *Hormones and Behavior*, *56*, 128–132. <https://doi.org/10.1016/j.yhbeh.2009.03.019>
- Thome, J., Liebke, L., Bungert, M., Schmahl, C., Domes, G., Bohus, M., & Lis, S. (2016). Confidence in facial emotion recognition in borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment*, *7*(2), 159–168. Scopus. <https://doi.org/10.1037/per0000142>
- Tolpin, L. H., Gunthert, K. C., Cohen, L. H., & O'Neill, S. C. (2004). Borderline personality features and instability of daily negative affect and self-esteem. *Journal Of Personality*, *72*(1), 111–137.
- Torgersen, S., Kringlen, E., & Cramer, V. (2001). The prevalence of personality disorders in a community sample. *Archives of General Psychiatry*, *58*(6), 590–596. <https://doi.org/10.1001/archpsyc.58.6.590>
- Trull, T. J., Ueda, D., Conforti, K., & Doan, B.-T. (1997). Borderline personality disorder features in nonclinical young adults: 2 Two-year outcome. *Journal of Abnormal Psychology*, *106*(2), 307–314. <https://doi.org/10.1037/0021-843X.106.2.307>
- Unoka, Z., Fogd, D., Seres, I., Kéri, S., & Csukly, G. (2015). Early maladaptive schema-related impairment and co-occurring current major depressive episode-related enhancement of mental state decoding ability in borderline personality disorder.

Journal of Personality Disorders, 29(2), 145–162.

https://doi.org/10.1521/pedi_2014_28_146

Unoka, Z., Seres, I., Áspán, N., Bódi, N., & Kéri, S. (2009). Trust game reveals restricted interpersonal transactions in patients with borderline personality disorder. *Journal of Personality Disorders*, 23(4), 399–409. <https://doi.org/10.1521/pedi.2009.23.4.399>

Van Heel, M., Luyten, P., De Meulemeester, C., Vanwalleghem, D., Vermote, R., & Lowyck, B. (2019). Mentalizing based on external features in borderline personality disorder compared with healthy controls: The role of attachment dimensions and childhood trauma. *Journal of Personality Disorders*, 33(6), 736–750.

https://doi.org/10.1521/pedi_2019_33_373

Vanwoerden, S., Leavitt, J., Gallagher, M. W., Temple, J. R., & Sharp, C. (2019). Dating violence victimization and borderline personality pathology: Temporal associations from late adolescence to early adulthood. *Personality Disorders: Theory, Research, and Treatment*, 10(2), 132–142. <https://doi.org/10.1037/per0000324>

Vaskinn, A., Antonsen, B. T., Fretland, R. A., Dziobek, I., Sundet, K., & Wilberg, T. (2015). Theory of mind in women with borderline personality disorder or schizophrenia: Differences in overall ability and error patterns. *Frontiers in Psychology*, 6.

<http://ezproxy.uow.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2015-43518-001&site=ehost-live> anja.vaskinn@psykologi.uio.no

Weinbrecht, A., Niedeggen, M., Roepke, S., & Renneberg, B. (2018). Feeling excluded no matter what? Bias in the processing of social participation in borderline personality disorder. *NeuroImage: Clinical*, 19, 343–350.

<https://doi.org/10.1016/j.nicl.2018.04.031>

Wiesenfeller, J., Flasbeck, V., Brown, E. C., & Brüne, M. (2020). Approach and avoidance behavior in female patients with borderline personality disorder. *Frontiers in*

Behavioral Neuroscience, 14, 588874–588874.

<https://doi.org/10.3389/fnbeh.2020.588874>

Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods*, 38(1), 174–180.

<https://doi.org/10.3758/BF03192765>

Wolff, S., Stiglmayr, C., Bretz, H. J., Lammers, C.-H., & Auckenthaler, A. (2007). Emotion identification and tension in female patients with borderline personality disorder.

British Journal of Clinical Psychology, 46(3), 347–360.

<https://doi.org/10.1348/014466507X173736>

Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects; of ability, conscientiousness, and goal orientation. *Journal of Applied*

Psychology, 89(2), 231–247. <https://doi.org/10.1037/0021-9010.89.2.231>

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema Therapy: A practitioner's guide*. The Guilford Press.

Zabihzadeh, A., Maleki, G., Richman, M. J., Hatami, A., Alimardani, Z., & Heidari, M.

(2017). Affective and cognitive theory of mind in borderline personality disorder: The role of comorbid depression. *Psychiatry Research*, 257, 144–149.

<https://doi.org/10.1016/j.psychres.2017.07.034>

Zanarini, M. C., Frankenburg, F. R., Reich, D. B., & Fitzmaurice, G. (2010). Time to attainment of recovery from borderline personality disorder and stability of recovery:

A 10-year prospective follow-up study. *Am J Psychiatry*, 167(6), 663–667.

<https://doi.org/10.1176/appi.ajp.2009.09081130>

Zanarini, M. C., Frankenburg, F. R., Reich, D. B., Wedig, M. M., Conkey, L. C., &

Fitzmaurice, G. M. (2015). The course of marriage/sustained cohabitation and

- parenthood among borderline patients followed prospectively for 16 years. *Journal Of Personality Disorders*, 29(1), 62–70. https://doi.org/10.1521/pedi_2014_28_147
- Zanarini, M. C., Gunderson, J. G., Frankenburg, F. R., & Chauncey, D. L. (1990). Discriminating borderline personality disorder from other axis II disorders. *The American Journal Of Psychiatry*, 147(2), 161–167.
- Zanarini, M. C., Vujanovic, A. A., Parachini, E. A., Boulanger, J. L., Frankenburg, F. R., & Hennen, J. (2003). A screening measure for BPD: The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). *Journal Of Personality Disorders*, 17, 568–573.
- Zegarra-Valdivia, J. A., & Chino Vilca, B. N. (2019). Social cognition and executive function in borderline personality disorder: Evidence of altered cognitive processes. *Salud Mental*, 42(1), 33–42. <https://doi.org/10.17711/SM.0185-3325.2019.005>
- Zinchenko, O., & Arsalidou, M. (2018). Brain responses to social norms: Meta-analyses of fMRI studies. *Human Brain Mapping*, 39(2), 955–970. <https://doi.org/10.1002/hbm.23895>